

APPLIED DIFFERENTIAL GEOMETRY



A Modern Introduction

Vladimir G Ivancevic • Tijana T Ivancevic

APPLIED
DIFFERENTIAL
GEOMETRY

A Modern Introduction

This page intentionally left blank

APPLIED DIFFERENTIAL GEOMETRY



A Modern Introduction

Vladimir G Ivancevic

Defence Science and Technology Organisation, Australia

Tijana T Ivancevic

The University of Adelaide, Australia

 **World Scientific**

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

APPLIED DIFFERENTIAL GEOMETRY

A Modern Introduction

Copyright © 2007 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN-13 978-981-270-614-0

ISBN-10 981-270-614-3

Printed in Singapore.

Dedicated to:

Nitya, Atma and Kali

This page intentionally left blank

Preface

Applied Differential Geometry: A Modern Introduction is a graduate-level monographic textbook. It is designed as a comprehensive introduction into methods and techniques of modern differential geometry with its various physical and non-physical applications. In some sense, it is a continuation of our previous book, *Natural Biodynamics* (World Scientific, 2006), which contains all the necessary background for comprehensive reading of the current book. While the previous book was focused on biodynamic applications, the core applications of the new book are in the realm of modern theoretical physics, mainly following its central line: *Einstein–Feynman–Witten*. Other applications include (among others): control theory, robotics, neurodynamics, psychodynamics and socio-economical dynamics.

The book has six chapters. Each chapter contains both ‘pure mathematics’ and related ‘applications’ labelled by the word ‘APPLICATION’.

The first chapter provides a soft (‘plain-English’) introduction into manifolds and related geometrical structures, for all the interested readers without the necessary background. As a ‘snap-shot’ illustration, at the end of the first chapter, a paradigm of generic differential-geometric modelling is given, which is supposed to fit all above-mentioned applications.

The second chapter gives technical preliminaries for development of the modern applied differential geometry. These preliminaries include: (i) classical geometrical objects – *tensors*, (ii) both classical and modern physical objects – *actions*, and modern geometrical objects – *functors*.

The third chapter develops modern *manifold geometry*, together with its main physical and non-physical applications. This chapter is a necessary background for comprehensive reading of the remaining chapters.

The fourth chapter develops modern *bundle geometry*, together with its main physical and non-physical applications.

The fifth chapter develops modern *jet bundle geometry*, together with its main applications in non-autonomous mechanics and field physics. All material in this chapter is based on the previous chapter.

The sixth chapter develops modern geometrical machinery of Feynman's *path integrals*, together with their various physical and non-physical applications. For most of this chapter, only the third chapter is a necessary background, assuming a basic understanding of quantum mechanics (as provided in the above-mentioned World Scientific book, *Natural Dynamics*).

The book contains both an extensive Index (which allows easy connections between related topics) and a number of cited references related to modern applied differential geometry.

Our approach to dynamics of complex systems is somewhat similar to the approach to mathematical physics used at the beginning of the 20th Century by the two leading mathematicians: David Hilbert and John von Neumann – the approach of combining mathematical rigor with conceptual clarity, or *geometrical intuition* that underpins the rigor.

The intended audience includes (but is not restricted to) theoretical and mathematical physicists; applied and pure mathematicians; control, robotics and mechatronics engineers; computer and neural scientists; mathematically strong chemists, biologists, psychologists, sociologists and economists – both in academia and industry.

Compared to all differential-geometric books published so far, *Applied Differential Geometry: A Modern Introduction* has much wider variety of both physical and non-physical applications. After comprehensive reading of this book, a reader should be able to both *read* and *write* journal papers in such diverse fields as superstring & topological quantum field theory, nonlinear dynamics & control, robotics, biomechanics, neurodynamics, psychodynamics and socio-economical dynamics.

V. Ivancevic

Defence Science & Technology Organisation, Australia
e-mail: Vladimir.Ivancevic@dsto.defence.gov.au

T. Ivancevic

School of Mathematics, The University of Adelaide
e-mail: Tijana.Ivancevic@adelaide.edu.au

Adelaide
May, 2006

Acknowledgments

The authors wish to thank Land Operations Division, Defence Science & Technology Organisation, Australia, for the support in developing the *Human Biodynamics Engine* (HBE) and all HBE-related text in this monograph.

Finally, we express our gratitude to the World Scientific Publishing Company, and especially to Ms. Zhang Ji and Mr. Rhaimie Wahap.

This page intentionally left blank

Glossary of Frequently Used Symbols

General

- ‘iff’ means ‘if and only if’;
- ‘r.h.s’ means ‘right hand side’; ‘l.h.s’ means ‘left hand side’;
- ODE means ordinary differential equation, while PDE means partial differential equation;
- *Einstein’s summation convention over repeated indices* (not necessarily one up and one down) *is assumed in the whole text*, unless explicitly stated otherwise.

Sets

- \mathbb{N} – natural numbers;
 \mathbb{Z} – integers;
 \mathbb{R} – real numbers;
 \mathbb{C} – complex numbers;
 \mathbb{H} – quaternions;
 \mathbb{K} – number field of real numbers, complex numbers, or quaternions.

Maps

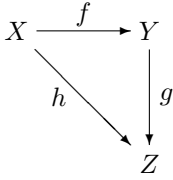
$f : A \rightarrow B$ – a function, (or map) between sets $A \equiv \text{Dom } f$ and $B \equiv \text{Cod } f$;

$\text{Ker } f = f^{-1}(e_B)$ – a kernel of f ;

$\text{Im } f = f(A)$ – an image of f ;

$\text{Coker } f = \text{Cod } f / \text{Im } f$ – a cokernel of f ;

$\text{Coim } f = \text{Dom } f / \text{Ker } f$ – a coimage of f ;



— a commutative diagram, requiring $h = g \circ f$.

Derivatives

$C^k(A, B)$ – set of k –times differentiable functions between sets A to B ;

$C^\infty(A, B)$ – set of *smooth* functions between sets A to B ;

$C^0(A, B)$ – set of *continuous* functions between sets A to B ;

$f'(x) = \frac{df(x)}{dx}$ – derivative of f with respect to x ;

\dot{x} – total time derivative of x ;

$\partial_t \equiv \frac{\partial}{\partial t}$ – partial time derivative;

$\partial_{x^i} \equiv \frac{\partial}{\partial x^i}$ – partial coordinate derivative;

$\dot{f} = \partial_t f + \partial_{x^i} f \dot{x}^i$ – total time derivative of the scalar field $f = f(t, x^i)$;

$u_t \equiv \partial_t u$, $u_x \equiv \partial_x u$, $u_{xx} \equiv \partial_{x^2} u$ – only in partial differential equations;

$L_{x^i} \equiv \partial_{x^i} L$, $L_{\dot{x}^i} \equiv \partial_{\dot{x}^i} L$ – coordinate and velocity partial derivatives of the Lagrangian function;

d – exterior derivative;

d^n – coboundary operator;

∂_n – boundary operator;

$\nabla = \nabla(g)$ – affine Levi–Civita connection on a smooth manifold M with Riemannian metric tensor $g = g_{ij}$;

Γ_{jk}^i – Christoffel symbols of the affine connection ∇ ;

$\nabla_X T$ – covariant derivative of the tensor–field T with respect to the vector–field X , defined by means of Γ_{jk}^i ;

$T_{;x^i} \equiv T_{|x^i}$ – covariant derivative of the tensor–field T with respect to the coordinate basis $\{x^i\}$;

$\ddot{T} \equiv \frac{DT}{dt} \equiv \frac{\nabla T}{dt}$ – absolute (intrinsic, or Bianchi) derivative of the tensor–field T upon the parameter t ; e.g., acceleration vector is the *absolute time derivative* of the velocity vector, $a^i = \dot{v}^i \equiv \frac{Dv^i}{dt}$; note that in general, $a^i \neq \dot{v}^i$ – this is crucial for *proper definition of Newtonian force*;

$\mathcal{L}_X T$ – Lie derivative of the tensor–field T in direction of the vector–field X ;

$[X, Y]$ – Lie bracket (commutator) of two vector–fields X and Y ;

$[F, G]$, or $\{F, G\}$ – Poisson bracket, or Lie–Poisson bracket, of two functions F and G .

Smooth Manifolds, Fibre Bundles and Jet Spaces

Unless otherwise specified, all *manifolds* M, N, \dots are assumed C^k -smooth, real, finite-dimensional, Hausdorff, paracompact, connected and without boundary,¹ while all maps are assumed C^k -smooth. We use the symbols \otimes, \vee, \wedge and \oplus for the tensor, symmetrized and exterior products, as well as the Whitney sum², respectively, while \lrcorner denotes the interior product (contraction) of (multi)vectors and p -forms, and \hookrightarrow denotes a manifold imbedding (i.e., both a submanifold and a topological subspace of the codomain manifold). The symbols ∂_B^A denote partial derivatives with respect to coordinates possessing multi-indices $_A^B$ (e.g., $\partial_\alpha = \partial/\partial x^\alpha$);

TM – tangent bundle of the manifold M ;

$\pi_M : TM \rightarrow M$ – natural projection;

T^*M – cotangent bundle of the manifold M ;

$\pi : Y \rightarrow X$ – fibre bundle;

(E, π, M) – vector bundle with total space E , base M and projection π ;

(Y, π, X, V) – fibre bundle with total space Y , base X , projection π and standard fibre V ;

$J^k(M, N)$ – space of k -jets of smooth functions between manifolds M and N ;

$J^k(X, Y)$ – k -jet space of a fibre bundle $Y \rightarrow X$; in particular, in mechanics we have a 1-jet space $J^1(\mathbb{R}, Q)$, with 1-jet coordinate maps $j_t^1 s : t \mapsto (t, x^i, \dot{x}^i)$, as well as a 2-jet space $J^2(\mathbb{R}, Q)$, with 2-jet coordinate maps $j_t^2 s : t \mapsto (t, x^i, \dot{x}^i, \ddot{x}^i)$;

$j_x^k s$ – k -jets of sections $s^i : X \rightarrow Y$ of a fibre bundle $Y \rightarrow X$;

We use the following kinds of *manifold maps*: immersion, imbedding, submersion, and projection. A map $f : M \rightarrow M'$ is called the *immersion* if the tangent map Tf at every point $x \in M$ is an injection (i.e., ‘1–1’ map). When f is both an immersion and an injection, its image is said to be a submanifold of M' . A submanifold which also is a topological subspace is called imbedded submanifold. A map $f : M \rightarrow M'$ is called *submersion* if the tangent map Tf at every point $x \in M$ is a surjection (i.e., ‘onto’ map). If f is both a submersion and a surjection, it is called *projection* or *fibre bundle*.

¹The only 1D manifolds obeying these conditions are the real line \mathbb{R} and the circle S^1 .

²Whitney sum \oplus is an analog of the direct (Cartesian) product for vector bundles. Given two vector bundles Y and Y' over the same base X , their Cartesian product is a vector bundle over $X \times X$. The *diagonal map* induces a vector bundle over X called the Whitney sum of these vector bundles and denoted by $Y \oplus Y'$.

Lie and (Co)Homology Groups

G – usually a general Lie group;

$GL(n)$ – general linear group with real coefficients in dimension n ;

$SO(n)$ – group of rotations in dimension n ;

T^n – toral (Abelian) group in dimension n ;

$Sp(n)$ – symplectic group in dimension n ;

$T(n)$ – group of translations in dimension n ;

$SE(n)$ – Euclidean group in dimension n ;

$H_n(M) = \text{Ker } \partial_n / \text{Im } \partial_{n-1}$ – n th homology group of the manifold M ;

$H^n(M) = \text{Ker } d^n / \text{Im } d^{n-1}$ – n th cohomology group of the manifold M .

Other Spaces and Operators

$i \equiv \sqrt{-1}$ – imaginary unit;

$C^k(M)$ – space of k -differentiable functions on the manifold M ;

$\Omega^k(M)$ – space of k -forms on the manifold M ;

\mathfrak{g} – Lie algebra of a Lie group G , i.e., the tangent space of G at its identity element;

$Ad(g)$ – adjoint endomorphism; recall that *adjoint representation* of a Lie group G is the linearized version of the action of G on itself by conjugation, i.e., for each $g \in G$, the inner automorphism $x \mapsto gxg^{-1}$ gives a linear transformation $Ad(g) : \mathfrak{g} \rightarrow \mathfrak{g}$, from the Lie algebra \mathfrak{g} of G to itself;

nD space (group, system) means n -dimensional space (group, system), for $n \in \mathbb{N}$;

\triangleright – semidirect (noncommutative) product; e.g., $SE(3) = SO(3) \triangleright \mathbb{R}^3$;

\lrcorner – interior product, or contraction, of a vector-field and a one-form;

\oint – Feynman path integral symbol, denoting integration over continuous spectrum of smooth paths and summation over discrete spectrum of Markov chains; e.g., $\oint \mathcal{D}[x] e^{iS[x]}$ denotes the path integral (i.e., sum-over-histories) over all possible paths $x^i = x^i(t)$ defined by the Hamilton action, $S[x] = \frac{1}{2} \int_{t_0}^{t_1} g_{ij} \dot{x}^i \dot{x}^j dt$, while $\oint \mathcal{D}[\Phi] e^{iS[\Phi]}$ denotes the path integral over all possible fields $\Phi^i = \Phi^i(x)$ defined by some field action $S[\Phi]$.

Categories

\mathcal{S} – all sets as objects and all functions between them as morphisms;

\mathcal{PS} – all pointed sets as objects and all functions between them preserving

base point as morphisms;

\mathcal{V} – all vector spaces as objects and all linear maps between them as morphisms;

\mathcal{B} – Banach spaces over \mathbb{R} as objects and bounded linear maps between them as morphisms;

\mathcal{G} – all groups as objects, all homomorphisms between them as morphisms;

\mathcal{A} – Abelian groups as objects, homomorphisms between them as morphisms;

\mathcal{AL} – all algebras (over a given number field \mathbb{K}) as objects, all their homomorphisms between them as morphisms;

\mathcal{T} – all topological spaces as objects, all continuous functions between them as morphisms;

\mathcal{PT} – pointed topological spaces as objects, continuous functions between them preserving base point as morphisms;

\mathcal{TG} – all topological groups as objects, their continuous homomorphisms as morphisms;

\mathcal{M} – all smooth manifolds as objects, all smooth maps between them as morphisms;

\mathcal{M}_n – n D manifolds as objects, their local diffeomorphisms as morphisms;

\mathcal{LG} – all Lie groups as objects, all smooth homomorphisms between them as morphisms;

\mathcal{LAL} – all Lie algebras (over a given field \mathbb{K}) as objects, all smooth homomorphisms between them as morphisms;

\mathcal{TB} – all tangent bundles as objects, all smooth tangent maps between them as morphisms;

$\mathcal{T}^*\mathcal{B}$ – all cotangent bundles as objects, all smooth cotangent maps between them as morphisms;

\mathcal{VB} – all smooth vector bundles as objects, all smooth homomorphisms between them as morphisms;

\mathcal{FB} – all smooth fibre bundles as objects, all smooth homomorphisms between them as morphisms;

Symplec – all symplectic manifolds (i.e., physical phase-spaces), all symplectic maps (i.e., canonical transformations) between them as morphisms;

Hilbert – all Hilbert spaces and all unitary operators as morphisms.

This page intentionally left blank

Contents

<i>Preface</i>	vii
<i>Glossary of Frequently Used Symbols</i>	xi
1. Introduction	1
1.1 Manifolds and Related Geometrical Structures	1
1.1.1 Geometrical Atlas	6
1.1.2 Topological Manifolds	8
1.1.2.1 Topological manifolds without boundary	10
1.1.2.2 Topological manifolds with boundary .	10
1.1.2.3 Properties of topological manifolds . . .	10
1.1.3 Differentiable Manifolds	12
1.1.4 Tangent and Cotangent Bundles of Manifolds . .	14
1.1.4.1 Tangent Bundle of a Smooth Manifold .	14
1.1.4.2 Cotangent Bundle of a Smooth Manifold	15
1.1.4.3 Fibre-, Tensor-, and Jet-Bundles . . .	15
1.1.5 Riemannian Manifolds: Configuration Spaces for Lagrangian Mechanics	16
1.1.5.1 Riemann Surfaces	17
1.1.5.2 Riemannian Geometry	19
1.1.5.3 APPLICATION: Lagrangian Mechanics .	22
1.1.5.4 Finsler manifolds	25
1.1.6 Symplectic Manifolds: Phase-Spaces for Hamiltonian Mechanics	25

1.1.7	Lie Groups	28
1.1.7.1	APPLICATION: Physical Examples of Lie Groups	30
1.1.8	APPLICATION: A Bird View on Modern Physics	31
1.1.8.1	Three Pillars of 20th Century Physics	31
1.1.8.2	String Theory in ‘Plain English’	33
1.2	APPLICATION: Paradigm of Differential–Geometric Modelling of Dynamical Systems	46
2.	Technical Preliminaries: Tensors, Actions and Functors	51
2.1	Tensors: Local Machinery of Differential Geometry	51
2.1.1	Transformation of Coordinates and Elementary Tensors	51
2.1.1.1	Transformation of Coordinates	52
2.1.1.2	Scalar Invariants	53
2.1.1.3	Vectors and Covectors	53
2.1.1.4	Second–Order Tensors	54
2.1.1.5	Higher–Order Tensors	56
2.1.1.6	Tensor Symmetry	57
2.1.2	Euclidean Tensors	58
2.1.2.1	Basis Vectors and the Metric Tensor in \mathbb{R}^n	58
2.1.2.2	Tensor Products in \mathbb{R}^n	59
2.1.3	Covariant Differentiation	60
2.1.3.1	Christoffel’s Symbols	60
2.1.3.2	Geodesics	61
2.1.3.3	Covariant Derivative	61
2.1.3.4	Covariant Form of Differential Operators	62
2.1.3.5	Absolute Derivative	63
2.1.3.6	3D Curve Geometry: Frenet–Serret Formulae	64
2.1.3.7	Mechanical Acceleration and Force	64
2.1.4	APPLICATION: Covariant Mechanics	65
2.1.4.1	Riemannian Curvature Tensor	70
2.1.4.2	Exterior Differential Forms	71
2.1.4.3	The Covariant Force Law	76
2.1.5	APPLICATION: Nonlinear Fluid Dynamics	78
2.1.5.1	Continuity Equation	78

2.1.5.2	Forces Acting on a Fluid	80
2.1.5.3	Constitutive and Dynamical Equations	81
2.1.5.4	Navier–Stokes Equations	82
2.2	Actions: The Core Machinery of Modern Physics	83
2.3	Functors: Global Machinery of Modern Mathematics	87
2.3.1	Maps	88
2.3.1.1	Notes from Set Theory	88
2.3.1.2	Notes From Calculus	89
2.3.1.3	Maps	89
2.3.1.4	Algebra of Maps	89
2.3.1.5	Compositions of Maps	90
2.3.1.6	The Chain Rule	90
2.3.1.7	Integration and Change of Variables	90
2.3.1.8	Notes from General Topology	91
2.3.1.9	Topological Space	92
2.3.1.10	Homotopy	93
2.3.1.11	Commutative Diagrams	95
2.3.1.12	Groups and Related Algebraic Structures	97
2.3.2	Categories	102
2.3.3	Functors	105
2.3.4	Natural Transformations	108
2.3.4.1	Compositions of Natural Transformations	109
2.3.4.2	Dinatural Transformations	109
2.3.5	Limits and Colimits	111
2.3.6	Adjunction	111
2.3.7	Abelian Categorical Algebra	113
2.3.8	n –Categories	116
2.3.8.1	Generalization of ‘Small’ Categories	117
2.3.8.2	Topological Structure of n –Categories	121
2.3.8.3	Homotopy Theory and Related n –Categories	121
2.3.8.4	Categorification	123
2.3.9	APPLICATION: n –Categorical Framework for Higher Gauge Fields	124
2.3.10	APPLICATION: Natural Geometrical Structures	128
2.3.11	Ultimate Conceptual Machines: Weak n –Categories	132

3.	Applied Manifold Geometry	137
3.1	Introduction	137
3.1.1	Intuition behind Einstein's Geometrodynamics	138
3.1.2	Einstein's Geometrodynamics in Brief	142
3.2	Intuition Behind the Manifold Concept	143
3.3	Definition of a Differentiable Manifold	145
3.4	Smooth Maps between Smooth Manifolds	147
3.4.1	Intuition behind Topological Invariants of Manifolds	148
3.5	(Co)Tangent Bundles of Smooth Manifolds	150
3.5.1	Tangent Bundle and Lagrangian Dynamics	150
3.5.1.1	Intuition behind a Tangent Bundle	150
3.5.1.2	Definition of a Tangent Bundle	150
3.5.2	Cotangent Bundle and Hamiltonian Dynamics	153
3.5.2.1	Definition of a Cotangent Bundle	153
3.5.3	APPLICATION: Command/Control in Human– Robot Interactions	154
3.5.4	APPLICATION: Generalized Bidirectional Associative Memory	157
3.6	Tensor Fields on Smooth Manifolds	163
3.6.1	Tensor Bundle	163
3.6.1.1	Pull-Back and Push-Forward	164
3.6.1.2	Dynamical Evolution and Flow	165
3.6.1.3	Vector-Fields and Their Flows	167
3.6.1.4	Vector-Fields on M	167
3.6.1.5	Integral Curves as Dynamical Trajectories	168
3.6.1.6	Dynamical Flows on M	172
3.6.1.7	Categories of ODEs	173
3.6.2	Differential Forms on Smooth Manifolds	174
3.6.2.1	1-Forms on M	174
3.6.2.2	k -Forms on M	176
3.6.2.3	Exterior Differential Systems	179
3.6.3	Exterior Derivative and (Co)Homology	180
3.6.3.1	Intuition behind Cohomology	182
3.6.3.2	Intuition behind Homology	183
3.6.3.3	De Rham Complex and Homotopy Operators	185

3.6.3.4	Stokes Theorem and de Rham Cohomology	186
3.6.3.5	Euler–Poincaré Characteristics of M	188
3.6.3.6	Duality of Chains and Forms on M	188
3.6.3.7	Hodge Star Operator and Harmonic Forms	190
3.7	Lie Derivatives on Smooth Manifolds	192
3.7.1	Lie Derivative Operating on Functions	192
3.7.2	Lie Derivative of Vector Fields	194
3.7.3	Time Derivative of the Evolution Operator	197
3.7.4	Lie Derivative of Differential Forms	197
3.7.5	Lie Derivative of Various Tensor Fields	198
3.7.6	APPLICATION: Lie–Derivative Neurodynamics	200
3.7.7	Lie Algebras	202
3.8	Lie Groups and Associated Lie Algebras	202
3.8.1	Definition of a Lie Group	203
3.8.2	Actions of Lie Groups on Smooth Manifolds	207
3.8.3	Basic Dynamical Lie Groups	210
3.8.3.1	Galilei Group	210
3.8.3.2	General Linear Group	211
3.8.4	APPLICATION: Lie Groups in Biodynamics	212
3.8.4.1	Lie Groups of Joint Rotations	212
3.8.4.2	Euclidean Groups of Total Joint Motions	216
3.8.4.3	Group Structure of Biodynamical Manifold	222
3.8.5	APPLICATION: Dynamical Games on $SE(n)$ –Groups	227
3.8.5.1	Configuration Models for Planar Vehicles	227
3.8.5.2	Two–Vehicles Conflict Resolution Manoeuvres	228
3.8.5.3	Symplectic Reduction and Dynamical Games on $SE(2)$	230
3.8.5.4	Nash Solutions for Multi–Vehicle Manoeuvres	233
3.8.6	Classical Lie Theory	235
3.8.6.1	Basic Tables of Lie Groups and their Lie Algebras	236

3.8.6.2	Representations of Lie groups	239
3.8.6.3	Root Systems and Dynkin Diagrams . .	240
3.8.6.4	Simple and Semisimple Lie Groups and Algebras	245
3.9	Lie Symmetries and Prolongations on Manifolds	247
3.9.1	Lie Symmetry Groups	247
3.9.1.1	Exponentiation of Vector Fields on M .	247
3.9.1.2	Lie Symmetry Groups and General DEs	249
3.9.2	Prolongations	250
3.9.2.1	Prolongations of Functions	250
3.9.2.2	Prolongations of Differential Equations	251
3.9.2.3	Prolongations of Group Actions	252
3.9.2.4	Prolongations of Vector Fields	253
3.9.2.5	General Prolongation Formula	254
3.9.3	Generalized Lie Symmetries	256
3.9.3.1	Noether Symmetries	257
3.9.4	APPLICATION: Biophysical PDEs	261
3.9.4.1	The Heat Equation	261
3.9.4.2	The Korteweg–De Vries Equation	262
3.9.5	Lie-Invariant Geometric Objects	262
3.9.5.1	Robot Kinematics	262
3.9.5.2	Maurer–Cartan 1-Forms	264
3.9.5.3	General Structure of Integrable One-Forms	265
3.9.5.4	Lax Integrable Dynamical Systems . . .	267
3.9.5.5	APPLICATION: Burgers Dynamical System	268
3.10	Riemannian Manifolds and Their Applications	271
3.10.1	Local Riemannian Geometry	271
3.10.1.1	Riemannian Metric on M	272
3.10.1.2	Geodesics on M	277
3.10.1.3	Riemannian Curvature on M	278
3.10.2	Global Riemannian Geometry	281
3.10.2.1	The Second Variation Formula	281
3.10.2.2	Gauss–Bonnet Formula	284
3.10.2.3	Ricci Flow on M	285
3.10.2.4	Structure Equations on M	287
3.10.3	APPLICATION: Autonomous Lagrangian Dynamics	289

3.10.3.1	Basis of Lagrangian Dynamics	289
3.10.3.2	Lagrange–Poincaré Dynamics	291
3.10.4	Core Application: Search for Quantum Gravity .	292
3.10.4.1	What is Quantum Gravity?	292
3.10.4.2	Main Approaches to Quantum Gravity .	293
3.10.4.3	Traditional Approaches to Quantum Gravity	300
3.10.4.4	New Approaches to Quantum Gravity .	304
3.10.4.5	Black Hole Entropy	310
3.10.5	Basics of Morse and (Co)Bordism Theories	311
3.10.5.1	Morse Theory on Smooth Manifolds . .	311
3.10.5.2	(Co)Bordism Theory on Smooth Manifolds	314
3.11	Finsler Manifolds and Their Applications	316
3.11.1	Definition of a Finsler Manifold	316
3.11.2	Energy Functional, Variations and Extrema . . .	317
3.11.3	APPLICATION: Finsler–Lagrangian Field Theory .	321
3.11.4	Riemann–Finsler Approach to Information Geometry	323
3.11.4.1	Model Specification and Parameter Estimation	323
3.11.4.2	Model Evaluation and Testing	324
3.11.4.3	Quantitative Criteria	324
3.11.4.4	Selection Among Different Models . . .	327
3.11.4.5	Riemannian Geometry of Minimum Description Length	330
3.11.4.6	Finsler Approach to Information Geometry	333
3.12	Symplectic Manifolds and Their Applications	335
3.12.1	Symplectic Algebra	335
3.12.2	Symplectic Geometry	336
3.12.3	APPLICATION: Autonomous Hamiltonian Mechanics	338
3.12.3.1	Basics of Hamiltonian Mechanics	338
3.12.3.2	Library of Basic Hamiltonian Systems .	351
3.12.3.3	Hamilton–Poisson Mechanics	361
3.12.3.4	Completely Integrable Hamiltonian Systems	363

3.12.3.5	Momentum Map and Symplectic Reduction	372
3.12.4	Multisymplectic Geometry	374
3.13	APPLICATION: Biodynamics–Robotics	375
3.13.1	Muscle–Driven Hamiltonian Biodynamics	376
3.13.2	Hamiltonian–Poisson Biodynamical Systems	379
3.13.3	Lie–Poisson Neurodynamics Classifier	383
3.13.4	Biodynamical Functors	384
3.13.4.1	The Covariant Force Functor	384
3.13.4.2	Lie–Lagrangian Biodynamical Functor	385
3.13.4.3	Lie–Hamiltonian Biodynamical Functor	391
3.13.5	Biodynamical Topology	401
3.13.5.1	(Co)Chain Complexes in Biodynamics	401
3.13.5.2	Morse Theory in Biodynamics	405
3.13.5.3	Hodge–De Rham Theory in Biodynamics	415
3.13.5.4	Lagrangian–Hamiltonian Duality in Biodynamics	419
3.14	Complex and Kähler Manifolds and Their Applications	428
3.14.1	Complex Metrics: Hermitian and Kähler	431
3.14.2	Calabi–Yau Manifolds	436
3.14.3	Special Lagrangian Submanifolds	437
3.14.4	Dolbeault Cohomology and Hodge Numbers	438
3.15	Conformal Killing–Riemannian Geometry	441
3.15.1	Conformal Killing Vector–Fields and Forms on M	442
3.15.2	Conformal Killing Tensors and Laplacian Symmetry	443
3.15.3	APPLICATION: Killing Vector and Tensor Fields in Mechanics	445
3.16	APPLICATION: Lax–Pair Tensors in Gravitation	448
3.16.1	Lax–Pair Tensors	450
3.16.2	Geometrization of the 3–Particle Open Toda Lattice	452
3.16.2.1	Tensorial Lax Representation	453
3.16.3	4D Generalizations	456
3.16.3.1	Case I	456
3.16.3.2	Case II	457
3.16.3.3	Energy–Momentum Tensors	457

3.17	Applied Unorthodox Geometries	458
3.17.1	Noncommutative Geometry	458
3.17.1.1	Moyal Product and Noncommutative Algebra	458
3.17.1.2	Noncommutative Space–Time Manifolds	459
3.17.1.3	Symmetries and Diffeomorphisms on Deformed Spaces	462
3.17.1.4	Deformed Diffeomorphisms	465
3.17.1.5	Noncommutative Space–Time Geometry	467
3.17.1.6	Star–Products and Expanded Einstein–Hilbert Action	470
3.17.2	Synthetic Differential Geometry	473
3.17.2.1	Distributions	474
3.17.2.2	Synthetic Calculus in Euclidean Spaces	476
3.17.2.3	Spheres and Balls as Distributions	478
3.17.2.4	Stokes Theorem for Unit Sphere	480
3.17.2.5	Time Derivatives of Expanding Spheres	481
3.17.2.6	The Wave Equation	482
4.	Applied Bundle Geometry	485
4.1	Intuition Behind a Fibre Bundle	485
4.2	Definition of a Fibre Bundle	486
4.3	Vector and Affine Bundles	491
4.3.1	The Second Vector Bundle of the Manifold M	495
4.3.2	The Natural Vector Bundle	496
4.3.3	Vertical Tangent and Cotangent Bundles	498
4.3.3.1	Tangent and Cotangent Bundles Revisited	498
4.3.4	Affine Bundles	500
4.4	APPLICATION: Semi–Riemannian Geometrical Mechanics	501
4.4.1	Vector–Fields and Connections	501
4.4.2	Hamiltonian Structures on the Tangent Bundle	503
4.5	K –Theory and Its Applications	508
4.5.1	Topological K –Theory	508
4.5.1.1	Bott Periodicity Theorem	509
4.5.2	Algebraic K –Theory	510
4.5.3	Chern Classes and Chern Character	511
4.5.4	Atiyah’s View on K –Theory	515
4.5.5	Atiyah–Singer Index Theorem	518

4.5.6	The Infinite-Order Case	520
4.5.7	Twisted K -Theory and the Verlinde Algebra . .	523
4.5.8	APPLICATION: K -Theory in String Theory . . .	526
4.5.8.1	Classification of Ramond-Ramond Fluxes	526
4.5.8.2	Classification of D -Branes	528
4.6	Principal Bundles	529
4.7	Distributions and Foliations on Manifolds	533
4.8	APPLICATION: Nonholonomic Mechanics	534
4.9	APPLICATION: Geometrical Nonlinear Control	537
4.9.1	Introduction to Geometrical Nonlinear Control . .	537
4.9.2	Feedback Linearization	539
4.9.3	Nonlinear Controllability	547
4.9.4	Geometrical Control of Mechanical Systems . . .	554
4.9.4.1	Abstract Control System	554
4.9.4.2	Global Controllability of Linear Control Systems	555
4.9.4.3	Local Controllability of Affine Control Systems	555
4.9.4.4	Lagrangian Control Systems	556
4.9.4.5	Lie-Adaptive Control	566
4.9.5	Hamiltonian Optimal Control and Maximum Principle	567
4.9.5.1	Hamiltonian Control Systems	567
4.9.5.2	Pontryagin's Maximum Principle	570
4.9.5.3	Affine Control Systems	571
4.9.6	Brain-Like Control Functor in Biodynamics . . .	573
4.9.6.1	Functor Control Machine	574
4.9.6.2	Spinal Control Level	576
4.9.6.3	Cerebellar Control Level	581
4.9.6.4	Cortical Control Level	584
4.9.6.5	Open Liouville Neurodynamics and Biodynamical Self-Similarity	587
4.9.7	Brain-Mind Functorial Machines	594
4.9.7.1	Neurodynamical 2-Functor	594
4.9.7.2	Solitary 'Thought Nets' and 'Emerging Mind'	597
4.9.8	Geometrodynamics of Human Crowd	602
4.9.8.1	Crowd Hypothesis	603

4.9.8.2	Geometrodynamics of Individual Agents	603
4.9.8.3	Collective Crowd Geometrodynamics	605
4.10	Multivector-Fields and Tangent-Valued Forms	606
4.11	APPLICATION: Geometrical Quantization	614
4.11.1	Quantization of Hamiltonian Mechanics	614
4.11.2	Quantization of Relativistic Hamiltonian Mechanics	617
4.12	Symplectic Structures on Fiber Bundles	624
4.12.1	Hamiltonian Bundles	625
4.12.1.1	Characterizing Hamiltonian Bundles	625
4.12.1.2	Hamiltonian Structures	626
4.12.1.3	Marked Hamiltonian Structures	630
4.12.1.4	Stability	632
4.12.1.5	Cohomological Splitting	632
4.12.1.6	Homological Action of $Ham(M)$ on M	634
4.12.1.7	General Symplectic Bundles	636
4.12.1.8	Existence of Hamiltonian Structures	637
4.12.1.9	Classification of Hamiltonian Structures	642
4.12.2	Properties of General Hamiltonian Bundles	645
4.12.2.1	Stability	645
4.12.2.2	Functorial Properties	648
4.12.2.3	Splitting of Rational Cohomology	650
4.12.2.4	Hamiltonian Bundles and Gromov-Witten Invariants	654
4.12.2.5	Homotopy Reasons for Splitting	659
4.12.2.6	Action of the Homology of (M) on $H_*(M)$	661
4.12.2.7	Cohomology of General Symplectic Bundles	664
4.13	Clifford Algebras, Spinors and Penrose Twistors	666
4.13.1	Clifford Algebras and Modules	666
4.13.1.1	The Exterior Algebra	669
4.13.1.2	The Spin Group	672
4.13.1.3	4D Case	672
4.13.2	Spinors	675
4.13.2.1	Basic Properties	675
4.13.2.2	4D Case	677
4.13.2.3	(Anti) Self Duality	681

4.13.2.4	Hermitian Structure on the Spinors . . .	686
4.13.2.5	Symplectic Structure on the Spinors . . .	689
4.13.3	Penrose Twistor Calculus	691
4.13.3.1	Penrose Index Formalism	691
4.13.3.2	Twistor Calculus	698
4.13.4	APPLICATION: Rovelli's Loop Quantum Gravity .	701
4.13.4.1	Introduction to Loop Quantum Gravity	701
4.13.4.2	Formalism of Loop Quantum Gravity .	708
4.13.4.3	Loop Algebra	709
4.13.4.4	Loop Quantum Gravity	711
4.13.4.5	Loop States and Spin Network States .	712
4.13.4.6	Diagrammatic Representation of the States	715
4.13.4.7	Quantum Operators	716
4.13.4.8	Loop v.s. Connection Representation .	717
4.14	APPLICATION: Seiberg–Witten Monopole Field Theory .	718
4.14.1	SUSY Formalism	721
4.14.1.1	$N = 2$ Supersymmetry	721
4.14.1.2	$N = 2$ Super–Action	721
4.14.1.3	Spontaneous Symmetry–Breaking . . .	723
4.14.1.4	Holomorphy and Duality	724
4.14.1.5	The SW Prepotential	724
4.14.2	Clifford Actions, Dirac Operators and Spinor Bundles	725
4.14.2.1	Clifford Algebras and Dirac Operators .	727
4.14.2.2	$Spin$ and $Spin_c$ Structures	729
4.14.2.3	Spinor Bundles	730
4.14.2.4	The Gauge Group and Its Equations . .	731
4.14.3	Original SW Low Energy Effective Field Action .	732
4.14.4	QED With Matter	735
4.14.5	QCD With Matter	737
4.14.6	Duality	738
4.14.6.1	Witten's Formalism	740
4.14.7	Structure of the Moduli Space	746
4.14.7.1	Singularity at Infinity	746
4.14.7.2	Singularities at Strong Coupling	747
4.14.7.3	Effects of a Massless Monopole	748
4.14.7.4	The Third Singularity	749

4.14.7.5	Monopole Condensation and Confinement	750
4.14.8	Masses and Periods	752
4.14.9	Residues	754
4.14.10	SW Monopole Equations and Donaldson Theory	757
4.14.10.1	Topological Invariance	760
4.14.10.2	Vanishing Theorems	762
4.14.10.3	Computation on Kähler Manifolds	765
4.14.11	SW Theory and Integrable Systems	769
4.14.11.1	$SU(N)$ Elliptic CM System	771
4.14.11.2	CM Systems Defined by Lie Algebras	772
4.14.11.3	Twisted CM–Systems Defined by Lie Algebras	773
4.14.11.4	Scaling Limits of CM–Systems	774
4.14.11.5	Lax Pairs for CM–Systems	776
4.14.11.6	CM and SW Theory for $SU(N)$	779
4.14.11.7	CM and SW Theory for General Lie Algebra	782
4.14.12	SW Theory and WDVV Equations	784
4.14.12.1	WDVV Equations	784
4.14.12.2	Perturbative SW Prepotentials	787
4.14.12.3	Associativity Conditions	789
4.14.12.4	SW Theories and Integrable Systems	790
4.14.12.5	WDVV Equations in SW Theories	793
5.	Applied Jet Geometry	797
5.1	Intuition Behind a Jet Space	797
5.2	Definition of a 1–Jet Space	801
5.3	Connections as Jet Fields	806
5.3.1	Principal Connections	815
5.4	Definition of a 2–Jet Space	818
5.5	Higher–Order Jet Spaces	822
5.6	APPLICATION: Jets and Non–Autonomous Dynamics	824
5.6.1	Geodesics	830
5.6.2	Quadratic Dynamical Equations	831
5.6.3	Equation of Free–Motion	832
5.6.4	Quadratic Lagrangian and Newtonian Systems	833
5.6.5	Jacobi Fields	835
5.6.6	Constraints	836

5.6.7	Time-Dependent Lagrangian Dynamics	841
5.6.8	Time-Dependent Hamiltonian Dynamics	843
5.6.9	Time-Dependent Constraints	848
5.6.10	Lagrangian Constraints	849
5.6.11	Quadratic Degenerate Lagrangian Systems	852
5.6.12	Time-Dependent Integrable Hamiltonian Systems	855
5.6.13	Time-Dependent Action-Angle Coordinates	858
5.6.14	Lyapunov Stability	860
5.6.15	First-Order Dynamical Equations	861
5.6.16	Lyapunov Tensor and Stability	863
	5.6.16.1 Lyapunov Tensor	863
	5.6.16.2 Lyapunov Stability	864
5.7	APPLICATION: Jets and Multi-Time Rheonomic Dynamics	868
	5.7.1 Relativistic Rheonomic Lagrangian Spaces	870
	5.7.2 Canonical Nonlinear Connections	871
	5.7.3 Cartan's Canonical Connections	874
	5.7.4 General Nonlinear Connections	876
5.8	Jets and Action Principles	877
5.9	APPLICATION: Jets and Lagrangian Field Theory	883
	5.9.1 Lagrangian Conservation Laws	888
	5.9.2 General Covariance Condition	893
5.10	APPLICATION: Jets and Hamiltonian Field Theory	897
	5.10.1 Covariant Hamiltonian Field Systems	899
	5.10.2 Associated Lagrangian and Hamiltonian Systems	902
	5.10.3 Evolution Operator	904
	5.10.4 Quadratic Degenerate Systems	910
5.11	APPLICATION: Gauge Fields on Principal Connections	913
	5.11.1 Connection Strength	913
	5.11.2 Associated Bundles	914
	5.11.3 Classical Gauge Fields	915
	5.11.4 Gauge Transformations	917
	5.11.5 Lagrangian Gauge Theory	919
	5.11.6 Hamiltonian Gauge Theory	920
	5.11.7 Gauge Conservation Laws	923
	5.11.8 Topological Gauge Theories	924
5.12	APPLICATION: Modern Geometrodynamics	928
	5.12.1 Stress-Energy-Momentum Tensors	928
	5.12.2 Gauge Systems of Gravity and Fermion Fields	955

5.12.3	Hawking–Penrose Quantum Gravity and Black Holes	963
6.	Geometrical Path Integrals and Their Applications	983
6.1	Intuition Behind a Path Integral	984
6.1.1	Classical Probability Concept	984
6.1.2	Discrete Random Variable	984
6.1.3	Continuous Random Variable	984
6.1.4	General Markov Stochastic Dynamics	985
6.1.5	Quantum Probability Concept	989
6.1.6	Quantum Coherent States	991
6.1.7	Dirac’s $\langle bra ket \rangle$ Transition Amplitude	992
6.1.8	Feynman’s Sum-over-Histories	994
6.1.9	The Basic Form of a Path Integral	996
6.1.10	APPLICATION: Adaptive Path Integral	997
6.2	Path Integral History	998
6.2.1	Extract from Feynman’s Nobel Lecture	998
6.2.2	Lagrangian Path Integral	1002
6.2.3	Hamiltonian Path Integral	1003
6.2.4	Feynman–Kac Formula	1004
6.2.5	Itô Formula	1006
6.3	Standard Path–Integral Quantization	1006
6.3.1	Canonical versus Path–Integral Quantization	1006
6.3.2	APPLICATION: Particles, Sources, Fields and Gauges	1011
6.3.2.1	Particles	1011
6.3.2.2	Sources	1012
6.3.2.3	Fields	1013
6.3.2.4	Gauges	1013
6.3.3	Riemannian–Symplectic Geometries	1014
6.3.4	Euclidean Stochastic Path Integral	1016
6.3.5	APPLICATION: Stochastic Optimal Control	1020
6.3.5.1	Path–Integral Formalism	1021
6.3.5.2	Monte Carlo Sampling	1023
6.3.6	APPLICATION: Nonlinear Dynamics of Option Pricing	1025
6.3.6.1	Theory and Simulations of Option Pricing	1025
6.3.6.2	Option Pricing via Path Integrals	1029

6.3.6.3	Continuum Limit and American Options	1035
6.3.7	APPLICATION: Nonlinear Dynamics of Complex Nets	1036
6.3.7.1	Continuum Limit of the Kuramoto Net	1037
6.3.7.2	Path–Integral Approach to Complex Nets	1038
6.3.8	APPLICATION: Dissipative Quantum Brain Model	1039
6.3.9	APPLICATION: Cerebellum as a Neural Path–Integral	1043
6.3.9.1	Spinal Autogenetic Reflex Control	1045
6.3.9.2	Cerebellum – the Comparator	1047
6.3.9.3	Hamiltonian Action and Neural Path Integral	1049
6.3.10	Path Integrals via Jets: Perturbative Quantum Fields	1050
6.4	Sum over Geometries and Topologies	1055
6.4.1	Simplicial Quantum Geometry	1057
6.4.2	Discrete Gravitational Path Integrals	1059
6.4.3	Regge Calculus	1061
6.4.4	Lorentzian Path Integral	1064
6.4.5	APPLICATION: Topological Phase Transitions and Hamiltonian Chaos	1069
6.4.5.1	Phase Transitions in Hamiltonian Systems	1069
6.4.5.2	Geometry of the Largest Lyapunov Exponent	1072
6.4.5.3	Euler Characteristics of Hamiltonian Systems	1075
6.4.6	APPLICATION: Force–Field Psychodynamics	1079
6.4.6.1	Motivational Cognition in the Life Space Foam	1080
6.5	APPLICATION: Witten’s TQFT, SW–Monopoles and Strings	1097
6.5.1	Topological Quantum Field Theory	1097
6.5.2	Seiberg–Witten Theory and TQFT	1103
6.5.2.1	SW Invariants and Monopole Equations	1103
6.5.2.2	Topological Lagrangian	1105
6.5.2.3	Quantum Field Theory	1107

6.5.2.4	Dimensional Reduction and 3D Field Theory	1112
6.5.2.5	Geometrical Interpretation	1115
6.5.3	TQFTs Associated with SW–Monopoles	1118
6.5.3.1	Dimensional Reduction	1122
6.5.3.2	TQFTs of 3D Monopoles	1124
6.5.3.3	Non–Abelian Case	1135
6.5.4	Stringy Actions and Amplitudes	1138
6.5.4.1	Strings	1139
6.5.4.2	Interactions	1140
6.5.4.3	Loop Expansion – Topology of Closed Surfaces	1141
6.5.5	Transition Amplitudes for Strings	1143
6.5.6	Weyl Invariance and Vertex Operator Formulation	1146
6.5.7	More General Actions	1146
6.5.8	Transition Amplitude for a Single Point Particle	1147
6.5.9	Witten’s Open String Field Theory	1148
6.5.9.1	Operator Formulation of String Field Theory	1149
6.5.9.2	Open Strings in Constant B –Field Background	1151
6.5.9.3	Construction of Overlap Vertices	1154
6.5.9.4	Transformation of String Fields	1164
6.6	APPLICATION: Dynamics of Strings and Branes	1168
6.6.1	A Relativistic Particle	1169
6.6.2	A String	1171
6.6.3	A Brane	1173
6.6.4	String Dynamics	1175
6.6.5	Brane Dynamics	1177
6.7	APPLICATION: Topological String Theory	1180
6.7.1	Quantum Geometry Framework	1180
6.7.2	Green–Schwarz Bosonic Strings and Branes	1181
6.7.3	Calabi–Yau Manifolds, Orbifolds and Mirror Symmetry	1186
6.7.4	More on Topological Field Theories	1189
6.7.5	Topological Strings	1204
6.7.6	Geometrical Transitions	1221
6.7.7	Topological Strings and Black Hole Attractors	1225

6.8	APPLICATION: Advanced Geometry and Topology of String Theory	1232
6.8.1	String Theory and Noncommutative Geometry . .	1232
6.8.1.1	Noncommutative Gauge Theory	1233
6.8.1.2	Open Strings in the Presence of Constant B -Field	1235
6.8.2	K -Theory Classification of Strings	1241
	<i>Bibliography</i>	1253
	<i>Index</i>	1295

Chapter 1

Introduction

In this introductory chapter we will firstly give a soft, ‘plain English’ introduction into manifolds and related differential–geometric terms, with intention to make this book accessible to the wider scientific and engineering community. Secondly, we will present the *paradigm* of differential–geometric modelling of dynamical systems, in the form of a generic algorithmic ‘recipe’ (see [Ivancevic and Ivancevic (2006)] for background details). The readers familiar with the manifold concept can skip the first section and only quickly review the second one.

1.1 Manifolds and Related Geometrical Structures

The core of both differential geometry and modern geometrical dynamics represents the concept of *manifold*. A *manifold* is an abstract mathematical space, which locally (i.e., in a close–up view) resembles the spaces described by *Euclidean geometry*, but which globally (i.e., when viewed as a whole) may have a more complicated structure. As main pure–mathematical references, we recommend popular graduate textbooks by two ex–*Bourbaki* members, *Serge Lang* [Lang (2005); Lang (2003); Lang (2002); Lang (1999)] and Jean Dieudonné [Dieudonné (1969); Dieudonné (1988)]. Besides, the reader might wish to consult some other ‘classics’, including [Spivak (1965); Spivak (1970–75); Choquet–Bruhat and DeWitt–Morete (1982); Bott and Tu (1982); Abraham *et al.* (1988); De Rham (1984); Milnor (1997); Munkres (1999)], as well as free internet sources [Wikipedia (2005); Weisstein (2004); PlanetMath (2006)]. Finally, as first–order applications, we recommend three popular textbooks in mechanics, [Abraham and Marsden (1978); Arnold (1989); Marsden and Ratiu (1999)].

For example, the *surface of Earth* is a manifold; locally it seems to be

flat, but viewed as a whole from the outer space (globally) it is actually round. A manifold can be constructed by ‘gluing’ separate *Euclidean spaces* together; for example, a world map can be made by gluing many maps of local regions together, and accounting for the resulting distortions.¹

Another example of a manifold is a *circle* S^1 . A small piece of a circle appears to be like a slightly-bent part of a straight line segment, but overall the circle and the segment are different 1D manifolds (see Figure 1.1). A circle can be formed by bending a straight line segment and gluing the ends together.²

¹On a sphere, the sum of the angles of a triangle is not equal to 180° . A sphere is not a Euclidean space, but locally the laws of the Euclidean geometry are good approximations. In a small triangle on the face of the earth, the sum of the angles is very nearly 180° . A sphere can be represented by a collection of two dimensional maps, therefore a sphere is a manifold.

²Locally, the circle looks like a line. It is 1D, that is, only one coordinate is needed to say where a point is on the circle locally. Consider, for instance, the top part of the circle (Figure 1.1), where the y -coordinate is positive. Any point in this part can be described by the x -coordinate. So, there is a continuous *bijection* χ_{top} (a mapping which is 1-1 both ways), which maps the top part of the circle to the open interval $(-1, 1)$, by simply projecting onto the first coordinate: $\chi_{top}(x, y) = x$. Such a function is called a *chart*. Similarly, there are charts for the bottom, left, and right parts of the circle. Together, these parts *cover* the whole circle and the four charts form an *atlas* (see the next subsection) for the circle. The top and right charts overlap: their intersection lies in the quarter of the circle where both the x - and the y -coordinates are positive. The two charts χ_{top} and χ_{right} map this part bijectively to the interval $(0, 1)$. Thus a function T from $(0, 1)$ to itself can be constructed, which first inverts the top chart to reach the circle and then follows the right chart back to the interval:

$$T(a) = \chi_{right} \left(\chi_{top}^{-1}(a) \right) = \chi_{right} \left(a, \sqrt{1 - a^2} \right) = \sqrt{1 - a^2}.$$

Such a function is called a *transition map*. The top, bottom, left, and right charts show that the circle is a manifold, but they do not form the only possible atlas. Charts need not be geometric projections, and the number of charts is a matter of choice. T and the other transition functions in Figure 1.1 are differentiable on the interval $(0, 1)$. Therefore, with this atlas the circle is a *differentiable*, or *smooth manifold*.

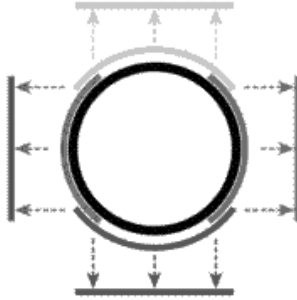


Fig. 1.1 The four charts each map part of the circle to an open interval, and together cover the whole circle.

The surfaces of a *sphere*³ and a *torus*⁴ are examples of 2D manifolds.

³The surface of the sphere S^2 can be treated in almost the same way as the circle S^1 . It can be viewed as a subset of \mathbb{R}^3 , defined by: $S = \{(x, y, z) \in \mathbb{R}^3 | x^2 + y^2 + z^2 = 1\}$. The sphere is 2D, so each chart will map part of the sphere to an open subset of \mathbb{R}^2 . Consider the northern hemisphere, which is the part with positive z coordinate. The function χ defined by $\chi(x, y, z) = (x, y)$, maps the northern hemisphere to the open unit disc by projecting it on the (x, y) -plane. A similar chart exists for the southern hemisphere. Together with two charts projecting on the (x, z) -plane and two charts projecting on the (y, z) -plane, an atlas of six charts is obtained which covers the entire sphere. This can be easily generalized to an n D sphere $S^n = \{(x_1, x_2, \dots, x_n) \in \mathbb{R}^n | x_1^2 + x_2^2 + \dots + x_n^2 = 1\}$.

An n -sphere S^n can be also constructed by gluing together two copies of \mathbb{R}^n . The transition map between them is defined as $\mathbb{R}^n \setminus \{0\} \rightarrow \mathbb{R}^n \setminus \{0\} : x \mapsto x/\|x\|^2$. This function is its own inverse, so it can be used in both directions. As the transition map is a (C^∞) -smooth function, this atlas defines a *smooth manifold*.

⁴A torus (pl. tori), denoted by T^2 , is a doughnut-shaped surface of revolution generated by revolving a circle about an axis coplanar with the circle. The sphere S^2 is a special case of the torus obtained when the axis of rotation is a diameter of the circle. If the axis of rotation does not intersect the circle, the torus has a hole in the middle and resembles a ring doughnut, a hula hoop or an inflated tire. The other case, when the axis of rotation is a chord of the circle, produces a sort of squashed sphere resembling a round cushion.

A torus can be defined parametrically by:

$$x(u, v) = (R + r \cos v) \cos u, \quad y(u, v) = (R + r \cos v) \sin u, \quad z(u, v) = r \sin v,$$

where $u, v \in [0, 2\pi]$, R is the distance from the center of the tube to the center of the torus, and r is the radius of the tube. According to a broader definition, the generator of a torus need not be a circle but could also be an ellipse or any other conic section.

Topologically, a torus is a closed surface defined as product of two circles: $T^2 = S^1 \times S^1$. The surface described above, given the relative topology from \mathbb{R}^3 , is *homeomorphic* to a topological torus as long as it does not intersect its own axis.

One can easily generalize the torus to arbitrary dimensions. An n -torus T^n is defined as a product of n circles: $T^n = S^1 \times S^1 \times \dots \times S^1$. Equivalently, the n -torus is obtained from the n -cube (the \mathbb{R}^n -generalization of the ordinary cube in \mathbb{R}^3) by gluing

Manifolds are important objects in mathematics, physics and control theory, because they allow more complicated structures to be expressed and understood in terms of the well-understood properties of simpler Euclidean spaces.

The *Cartesian product* of manifolds is also a manifold (note that not every manifold can be written as a product). The dimension of the product manifold is the sum of the dimensions of its factors. Its topology is the product topology, and a Cartesian product of charts is a chart for the product manifold. Thus, an atlas for the product manifold can be constructed using atlases for its factors. If these atlases define a differential structure on the factors, the corresponding atlas defines a differential structure on the product manifold. The same is true for any other structure defined on the factors. If one of the factors has a boundary, the product manifold also has a boundary. Cartesian products may be used to construct tori and cylinders, for example, as $S^1 \times S^1$ and $S^1 \times [0, 1]$, respectively.

Manifolds need not be *connected* (all in ‘one piece’): a pair of separate circles is also a *topological manifold* (see below). Manifolds need not be *closed*: a line segment without its ends is a manifold. Manifolds need not be *finite*: a parabola is a topological manifold.

Manifolds can be viewed using either extrinsic or intrinsic view. In the *extrinsic view*, usually used in geometry and topology of surfaces, an n D manifold M is seen as embedded in an $(n + 1)$ D Euclidean space \mathbb{R}^{n+1} . Such a manifold is called a ‘codimension 1 space’. With this view it is easy to use intuition from Euclidean spaces to define additional structure. For example, in a Euclidean space it is always clear whether a vector at some point is tangential or normal to some surface through that point. On the other hand, the *intrinsic view* of an n D manifold M is an abstract way of considering M as a topological space by itself, without any need for surrounding $(n + 1)$ D Euclidean space. This view is more flexible and thus it is usually used in high-dimensional mechanics and physics (where manifolds used represent configuration and phase spaces of dynamical systems), can make it harder to imagine what a tangent vector might be.

Additional structures are often defined on manifolds. Examples of manifolds with additional structure include:

- *differentiable* (or, *smooth manifolds*, on which one can do calculus;

the opposite faces together.

An n -torus T^n is an example of an n D *compact manifold*. It is also an important example of a *Lie group* (see below).

- *Riemannian manifolds*, on which *distances* and *angles* can be defined;
- *symplectic manifolds*, which serve as the *phase space* in mechanics and physics;
- 4D pseudo-Riemannian manifolds which model *space-time* in general relativity.

The study of manifolds combines many important areas of mathematics: it generalizes concepts such as curves and surfaces as well as ideas from linear algebra and topology. Certain special classes of manifolds also have additional algebraic structure; they may behave like groups, for instance.

Historically, before the modern concept of a manifold there were several important results:

- (1) *Carl Friedrich Gauss* was arguably the first to consider abstract spaces as mathematical objects in their own right. His ‘Theorema Egregium’ gives a method for computing the curvature of a surface S without considering the ambient *Euclidean 3D space* \mathbb{R}^3 in which the surface lies. Such a surface would, in modern terminology, be called a manifold.
- (2) *Non-Euclidean geometry* considers spaces where Euclid’s ‘Parallel Postulate’ fails. Saccheri first studied them in 1733. Lobachevsky, Bolyai, and Riemann developed them 100 years later. Their research uncovered two more types of spaces whose geometric structures differ from that of classical *Euclidean n D space* \mathbb{R}^n ; these gave rise to *hyperbolic geometry* and *elliptic geometry*. In the modern theory of manifolds, these notions correspond to manifolds with negative and positive *curvature*, respectively.
- (3) The *Euler characteristic* is an example of a *topological property* (or *topological invariant*) of a manifold. For a convex polyhedron in Euclidean 3D space \mathbb{R}^3 , with V vertices, E edges and F faces, Euler showed that $V - E + F = 2$. Thus the number 2 is called the Euler characteristic of the space \mathbb{R}^3 . The Euler characteristic of other 3D spaces is a useful topological invariant, which can be extended to higher dimensions using the so-called *Betti numbers*. The study of other topological invariants of manifolds is one of the central themes of topology.
- (4) *Bernhard Riemann* was the first to do extensive work generalizing the idea of a surface to higher dimensions. The name manifold comes from Riemann’s original German term, ‘Mannigfaltigkeit’, which W.K. Clifford translated as ‘*manifoldness*’. In his famous Göttingen inaugural

lecture entitled ‘On the Hypotheses which lie at the Bases of Geometry’, Riemann described the set of all possible values of a variable with certain constraints as a ‘manifoldness’, because the variable can have many values. He distinguishes between continuous manifoldness and discontinuous manifoldness, depending on whether the value changes continuously or not. As continuous examples, Riemann refers to not only colors and the locations of objects in space, but also the possible shapes of a spatial figure. Using *mathematical induction*, Riemann constructs an n times extended manifoldness, or n D manifoldness, as a continuous stack of $(n - 1)$ D manifoldnesses. Riemann’s intuitive notion of a ‘manifoldness’ evolved into what is today formalized as a manifold.

- (5) *Henri Poincaré* studied 3D manifolds at the end of the 19th Century, and raised a question, today known as the *Poincaré conjecture*. *Hermann Weyl* gave an *intrinsic definition for differentiable manifolds* in 1912. During the 1930s, *H. Whitney* and others clarified the foundational aspects of the subject, and thus intuitions dating back to the latter half of the 19th Century became precise, and developed through differential geometry (in particular, by the *Lie group theory* introduced by *Sophus Lie* in 1870, see below).

1.1.1 Geometrical Atlas

Now we continue introducing manifolds. As already stated above, an *atlas* describes how a complicated space called a manifold is glued together from simpler pieces. Each piece is given by a *chart* (also known as *coordinate chart* or *local coordinate system*).⁵

The description of most manifolds requires more than one chart (a single chart is adequate for only the simplest manifolds). An atlas is a specific collection of charts which covers a manifold. An atlas is not unique as all manifolds can be covered multiple ways using different combinations of

⁵A coordinate map, a coordinate chart, or simply a chart, of a manifold is an invertible map between a subset of the manifold and a simple space such that both the map and its inverse preserve the desired structure. For a *topological manifold*, the simple space is some Euclidean space \mathbb{R}^n and interest is focused on the *topological structure*. This structure is preserved by *homeomorphisms* (invertible maps that are continuous in both directions).

In the case of a *differentiable manifold*, a set of charts called an atlas allows us to do calculus on manifolds. Polar coordinates, for example, form a chart for the plane \mathbb{R}^2 minus the positive x -axis and the origin. Another example of a chart is the map χ_{top} mentioned above, a chart for the circle.

charts.

The atlas containing all possible charts consistent with a given atlas is called the *maximal atlas*. Unlike an ordinary atlas, the maximal atlas of a given atlas is unique.

More generally, an atlas for a complicated space is constructed out of the following pieces of information:

- (i) A list of spaces that are considered simple.
- (ii) For each point in the complicated space, a neighborhood of that point that is homeomorphic to a simple space, the homeomorphism being a chart.

(iii) We require the different charts to be *compatible*. At the minimum, we require that the composite of one chart with the inverse of another be a *homeomorphism* (also known as a *change of coordinates*, or a *transformation of coordinates*, or a *transition function*, or a *transition map*) but we usually impose stronger requirements, such as C^∞ -smoothness.⁶

This definition of atlas is exactly analogous to the non-mathematical meaning of atlas. Each individual map in an atlas of the world gives a neighborhood of each point on the globe that is homeomorphic to the plane. While each individual map does not exactly line up with other maps that it overlaps with (because of the Earth's curvature), the overlap of two maps can still be compared (by using latitude and longitude lines, for example).

Different choices for simple spaces and compatibility conditions give different objects. For example, if we choose for our simple spaces the *Euclidean spaces* \mathbb{R}^n , we get topological manifolds. If we also require the coordinate

⁶Charts in an atlas may *overlap* and a single point of a manifold may be represented in several charts. If two charts overlap, parts of them represent the same region of the manifold, just as a map of Europe and a map of Asia may both contain Moscow. Given two overlapping charts, a *transition function* can be defined, which goes from an open Euclidean n D ball $B^n = \{(x_1, x_2, \dots, x_n) \in \mathbb{R}^n | x_1^2 + x_2^2 + \dots + x_n^2 < 1\}$ in \mathbb{R}^n to the manifold and then back to another (or perhaps the same) open n D ball in \mathbb{R}^n . The resultant map, like the map T in the circle example above, is called a change of coordinates, a coordinate transformation, a transition function, or a transition map.

An atlas can also be used to define *additional structure on the manifold*. The structure is first defined on each chart separately. If all the transition maps are compatible with this structure, the structure transfers to the manifold.

This is the standard way differentiable manifolds are defined. If the transition functions of an atlas for a topological manifold preserve the natural differential structure of \mathbb{R}^n (that is, if they are *diffeomorphisms*, i.e., invertible maps that are smooth in both directions), the differential structure transfers to the manifold and turns it into a differentiable, or smooth manifold.

In general the structure on the manifold depends on the atlas, but sometimes different atlases give rise to the same structure. Such atlases are called *compatible*.

changes to be diffeomorphisms, we get *differentiable manifolds*, or *smooth manifolds*.

We call two atlases compatible if the charts in the two atlases are all compatible (or equivalently if the union of the two atlases is an atlas). Usually, we want to consider two compatible atlases as giving rise to the same space. Formally, (as long as our concept of compatibility for charts has certain simple properties), we can define an equivalence relation on the set of all atlases, calling two the same if they are compatible. In fact, the union of all atlases compatible with a given atlas is itself an atlas, called a complete (or maximal) atlas. Thus every atlas is contained in a unique complete atlas.

By definition, a smooth differentiable structure (or differential structure) on a manifold M is such a maximal atlas of charts, all related by smooth coordinate changes on the overlaps.

1.1.2 Topological Manifolds

A *topological manifold* is a *manifold* that is glued together from *Euclidean spaces* \mathbb{R}^n . Euclidean spaces are the simplest examples of topological manifolds. Thus, a topological manifold is a topological space that locally looks like an Euclidean space. More precisely, a topological manifold is a *topological space*⁷ locally homeomorphic to a Euclidean space. This means that

⁷Topological spaces are structures that allow one to formalize concepts such as *convergence*, *connectedness* and *continuity*. They appear in virtually every branch of modern mathematics and are a central unifying notion. Technically, a *topological space* is a set X together with a collection T of subsets of X satisfying the following axioms:

- (1) The empty set and X are in T ;
- (2) The union of any collection of sets in T is also in T ; and
- (3) The intersection of any pair of sets in T is also in T .

The collection T is a topology on X . The sets in T are the open sets, and their complements in X are the closed sets. The elements of X are called points. By induction, the intersection of any finite collection of open sets is open. Thus, the third Axiom can be replaced by the equivalent one that the topology be closed under all finite intersections instead of just pairwise intersections. This has the benefit that we need not explicitly require that X be in T , since the empty intersection is (by convention) X . Similarly, we can conclude that the empty set is in T by using Axiom 2. and taking a union over the empty collection. Nevertheless, it is conventional to include the first Axiom even when it is redundant.

A function between topological spaces is said to be continuous iff the inverse image of every open set is open. This is an attempt to capture the intuition that there are no ‘breaks’ or ‘separations’ in the function. A *homeomorphism* is a bijection that is continuous and whose inverse is also continuous. Two spaces are said to be homeomor-

every point has a neighborhood for which there exists a *homeomorphism* (a bijective continuous function whose inverse is also continuous) mapping that neighborhood to \mathbb{R}^n . These homeomorphisms are the charts of the manifold.

Usually additional technical assumptions on the topological space are made to exclude pathological cases. It is customary to require that the space be Hausdorff and second countable.

The *dimension* of the manifold at a certain point is the dimension of the Euclidean space charts at that point map to (number n in the definition). All points in a connected manifold have the same dimension.

In topology and related branches of mathematics, a *connected space* is a topological space which cannot be written as the disjoint union of two or more nonempty spaces. Connectedness is one of the principal topological properties that is used to distinguish topological spaces. A stronger notion is that of a *path-connected space*, which is a space where any two points can be joined by a *path*.⁸

phic if there exists a homeomorphism between them. From the standpoint of topology, homeomorphic spaces are essentially identical.

The category of topological spaces, *Top*, with topological spaces as objects and continuous functions as morphisms is one of the fundamental categories in mathematics. The attempt to classify the objects of this category (up to homeomorphism) by invariants has motivated and generated entire areas of research, such as homotopy theory, homology theory, and K-theory.

⁸Formally, for a topological space X the following conditions are equivalent:

- (1) X is connected.
- (2) X cannot be divided into two disjoint nonempty closed sets (this follows since the complement of an open set is closed).
- (3) The only sets which are both open and closed (open sets) are X and the empty set.
- (4) The only sets with empty boundary are X and the empty set.
- (5) X cannot be written as the union of two nonempty separated sets.

The maximal nonempty connected subsets of any topological space are called the *connected components* of the space. The components form a *partition* of the space (that is, they are disjoint and their union is the whole space). Every component is a closed subset of the original space. The components in general need not be open: the components of the rational numbers, for instance, are the one-point sets. A space in which all components are one-point sets is called *totally disconnected*.

The space X is said to be *path-connected* iff for any two points $x, y \in X$ there exists a *continuous function* $f : [0, 1] \rightarrow X$, from the unit interval $[0, 1]$ to X , with $f(0) = x$ and $f(1) = y$ (this function is called a *path* from x to y). Every path-connected space is connected.

1.1.2.1 Topological manifolds without boundary

The prototypical example of a topological manifold without boundary is Euclidean space. A general manifold without boundary looks locally, as a topological space, like Euclidean space. This is formalized by requiring that a manifold without boundary is a non-empty topological space in which every point has an open neighborhood homeomorphic to (an open subset of) \mathbb{R}^n (Euclidean n -space). Another way of saying this, using charts, is that a manifold without boundary is a non-empty topological space in which at every point there is an \mathbb{R}^n -chart.

1.1.2.2 Topological manifolds with boundary

Generally speaking, it is possible to allow a topological manifold to have a boundary. The prototypical example of a topological manifold with boundary is the Euclidean closed half-space. Most points in Euclidean closed half-space, those not on the boundary, have a neighborhood homeomorphic to Euclidean space in addition to having a neighborhood homeomorphic to Euclidean closed half-space, but the points on the boundary only have neighborhoods homeomorphic to Euclidean closed half-space and not to Euclidean space. Thus we need to allow for two kinds of points in our topological manifold with boundary: points in the interior and points in the boundary. Points in the interior will, as before, have neighborhoods homeomorphic to Euclidean space, but may also have neighborhoods homeomorphic to Euclidean closed half-space. Points in the boundary will have neighborhoods homeomorphic to Euclidean closed half-space. Thus a topological manifold with boundary is a non-empty topological space in which at each point there is an \mathbb{R}^n -chart or an $[0, \infty) \times \mathbb{R}^{n-1}$ -chart. The set of points at which there are only $[0, \infty) \times \mathbb{R}^{n-1}$ -charts is called the boundary and its complement is called the interior. The interior is always non-empty and is a topological n -manifold without boundary. If the boundary is non-empty then it is a topological $(n - 1)$ -manifold without boundary. If the boundary is empty, then we regain the definition of a topological manifold without boundary.

1.1.2.3 Properties of topological manifolds

A manifold with empty boundary is said to be *closed manifold* if it is compact, and *open manifold* if it is not compact.

All 1-manifolds are *curves* and all 2-manifolds are *surfaces*. Examples

of curves include circles, hyperbolas, and the trefoil knot. Sphere, cylinder, torus, projective plane,⁹ Möbius strip,¹⁰ and Klein bottle¹¹ are examples of surfaces.

Manifolds inherit many of the local properties of Euclidean space. In particular, they are locally path-connected, locally compact and locally metrizable. Being locally compact Hausdorff spaces, they are necessarily Tychonoff spaces. Requiring a manifold to be Hausdorff may seem strange; it is tempting to think that being locally homeomorphic to a Euclidean space implies being a Hausdorff space. A counterexample is created by deleting zero from the real line and replacing it with two points, an open neighborhood of either of which includes all nonzero numbers in some open interval centered at zero. This construction, called the real line with two origins is not Hausdorff, because the two origins cannot be separated.

All compact surfaces are homeomorphic to exactly one of the 2-sphere, a connected sum of tori, or a connected sum of projective planes.

A topological space is said to be homogeneous if its homeomorphism group acts transitively on it. Every connected manifold without boundary is homogeneous, but manifolds with nonempty boundary are not homogeneous.

It can be shown that a manifold is metrizable if and only if it is paracompact. Non-paracompact manifolds (such as the long line) are generally

⁹Begin with a sphere centered on the origin. Every line through the origin pierces the sphere in two opposite points called antipodes. Although there is no way to do so physically, it is possible to mathematically merge each antipode pair into a single point. The closed surface so produced is the real projective plane, yet another non-orientable surface. It has a number of equivalent descriptions and constructions, but this route explains its name: all the points on any given line through the origin projects to the same 'point' on this 1plane'.

¹⁰Begin with an infinite circular cylinder standing vertically, a manifold without boundary. Slice across it high and low to produce two circular boundaries, and the cylindrical strip between them. This is an orientable manifold with boundary, upon which 'surgery' will be performed. Slice the strip open, so that it could unroll to become a rectangle, but keep a grasp on the cut ends. Twist one end 180 deg, making the inner surface face out, and glue the ends back together seamlessly. This results in a strip with a permanent half-twist: the Möbius strip. Its boundary is no longer a pair of circles, but (topologically) a single circle; and what was once its 'inside' has merged with its 'outside', so that it now has only a single side.

¹¹Take two Möbius strips; each has a single loop as a boundary. Straighten out those loops into circles, and let the strips distort into cross-caps. Gluing the circles together will produce a new, closed manifold without boundary, the Klein bottle. Closing the surface does nothing to improve the lack of orientability, it merely removes the boundary. Thus, the Klein bottle is a closed surface with no distinction between inside and outside. Note that in 3D space, a Klein bottle's surface must pass through itself. Building a Klein bottle which is not self-intersecting requires four or more dimensions of space.

regarded as pathological, so it's common to add paracompactness to the definition of an n -manifold. Sometimes n -manifolds are defined to be second-countable, which is precisely the condition required to ensure that the manifold embeds in some finite-dimensional Euclidean space. Note that every compact manifold is second-countable, and every second-countable manifold is paracompact.

Topological manifolds are usually required to be Hausdorff and second-countable. Every Hausdorff, second countable manifold of dimension n admits an atlas consisting of at most $n + 1$ charts.

1.1.3 Differentiable Manifolds

For most applications, a special kind of topological manifold, a *differentiable manifold*, is used. If the local charts on a manifold are compatible in a certain sense, one can define directions, tangent spaces, and differentiable functions on that manifold. In particular it is possible to use calculus on a differentiable manifold. Each point of an n D differentiable manifold has a tangent space. This is an Euclidean space \mathbb{R}^n consisting of the tangent vectors of the curves through the point.

Two important classes of differentiable manifolds are *smooth* and *analytic* manifolds. For smooth manifolds the transition maps are smooth, that is infinitely differentiable, denoted by C^∞ . Analytic manifolds are smooth manifolds with the additional condition that the transition maps are analytic (a technical definition which loosely means that Taylor's expansion Theorem¹² holds). The sphere can be given analytic structure, as can most familiar curves and surfaces.

In other words, a differentiable (or, smooth) manifold is a topological manifold with a globally defined differentiable (or, smooth) structure. A topological manifold can be given a differentiable structure locally by using the homeomorphisms in the atlas of the topological space (i.e., the homeomorphism can be used to give a local coordinate system). The global differentiable structure is induced when it can be shown that the natural composition of the homeomorphisms between the corresponding open Euclidean spaces are differentiable on overlaps of charts in the atlas. Therefore, the coordinates defined by the homeomorphisms are differentiable with respect to each other when treated as real valued functions with respect to

¹²The most basic example of *Taylor's Theorem* is the approximation of the exponential function near the origin point $x = 0$: $e^x \approx 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}$. For technical details, see any calculus textbook.

the variables defined by other coordinate systems whenever charts overlap. This idea is often presented formally using transition maps.

This allows one to extend the meaning of *differentiability* to spaces without global coordinate systems. Specifically, a differentiable structure allows one to define a global differentiable tangent space, and consequently, differentiable functions, and differentiable tensor-fields (including vector-fields). Differentiable manifolds are very important in physics. Special kinds of differentiable manifolds form the arena for physical theories such as classical mechanics (*Hamiltonian mechanics* and *Lagrangian mechanics*), general relativity and *Yang–Mills gauge theory*. It is possible to develop calculus on differentiable manifolds, leading to such mathematical machinery as the exterior calculus.

Historically, the development of differentiable manifolds (as well as differential geometry in general) is usually credited to C.F. Gauss and his student B. Riemann. The work of physicists J.C. Maxwell and A. Einstein lead to the development of the theory transformations between coordinate systems which preserved the essential geometric properties. Eventually these ideas were generalized by H. Weyl in ‘Idee der Riemannschen Fläche’ (1913) and ‘Raum, Zeit, Materie’ (‘Space Time Matter’, 1921). T. Levi-Civita applied these ideas in ‘Lezioni di calcolo differenziale assoluto’ (‘The Absolute Differential Calculus’, 1923). The approach of Weyl was essentially to consider the coordinate functions in terms of other coordinates and to assume differentiability for the coordinate function. In 1963, S. Kobayashi and K. Nomizu gave the group transformation/atlas approach.

Generalizations of manifolds

The three most common generalizations of manifolds are:

- *orbifolds*: An orbifold is a generalization of manifold allowing for certain kinds of ‘singularities’ in the topology. Roughly speaking, it is a space which locally looks like the quotients of some simple space (e.g., Euclidean space) by the actions of various finite groups. The singularities correspond to fixed points of the group actions, and the actions must be compatible in a certain sense.
- *algebraic varieties and schemes*: An algebraic variety is glued together from affine algebraic varieties, which are zero sets of polynomials over algebraically closed fields. Schemes are likewise glued together from affine schemes, which are a generalization of algebraic varieties. Both

are related to manifolds, but are constructed using *sheaves*¹³ instead of atlases. Because of singular points one cannot assume a variety is a manifold.

- *CW-complexes*: A CW complex is a topological space formed by gluing objects of different dimensionality together; for this reason they generally are not manifolds. However, they are of central interest in algebraic topology, especially in homotopy theory, where such dimensional defects are acceptable.

1.1.4 Tangent and Cotangent Bundles of Manifolds

1.1.4.1 Tangent Bundle of a Smooth Manifold

The *tangent bundle* of an open contractable set $U \in \mathbb{R}^n$ is the smooth manifold $U \times \mathbb{R}^n$.

The tangent bundle TM of the smooth manifold M is constructed using the transition maps which define the differentiable structure of M . One may construct transition maps for the atlas of smooth manifolds $U_i \times \mathbb{R}^n$, where

¹³A sheaf F on a topological space X is an object that assigns a structure $F(U)$ (such as a set, group, or ring) to each open subset U of X . The structures $F(U)$ are compatible with the operations of restricting the open set to smaller subsets and gluing smaller open sets to get a bigger one. A presheaf is similar to a sheaf, but it may not be possible to glue. Sheaves enable one to discuss in a refined way what is a local property, as applied to a function.

Sheaves are used in topology, algebraic geometry and differential geometry whenever one wants to keep track of algebraic data that vary with every open set of the given geometrical space. They are a *global tool* to study objects which *vary locally* (that is, depend on the open sets). As such, they are a natural instrument to study the *global behavior* of objects which are of local nature, such as open sets, analytic functions, manifolds, and so on.

For a typical example, consider a topological space X , and for every open set $U \in X$, let $F(U)$ be the set of all continuous functions $U \rightarrow \mathbb{R}$. If V is an open subset of U , then the functions on U can be restricted to V , and we get a map $F(U) \rightarrow F(V)$. Now, ‘gluing’ describes the following process: suppose the U_i are given open sets with union U , and for each i we are given an element $f_i \in F(U_i)$, a continuous function $f_i : U_i \rightarrow \mathbb{R}$. If these functions agree where they overlap, then we can glue them together in a unique way to form a continuous function $f : U \rightarrow \mathbb{R}$, which agrees with all the given f_i . The collection of the sets $F(U)$ together with the restriction maps $F(U) \rightarrow F(V)$ then form a sheaf of sets on X . Furthermore, each $F(U)$ is a *commutative ring* and the restriction maps are *ring homomorphisms*, making F a *sheaf of rings* on X .

For a very similar example, consider a smooth manifold M , and for every open set U of M , let $F(U)$ be the set of smooth functions $U \rightarrow \mathbb{R}$. Here too, ‘gluing’ works and we get a sheaf of rings on M . Another sheaf on M assigns to every open set U of M the vector space of all smooth vector-fields defined on U . Restriction and gluing of vector-fields works like that of functions, and we get a *sheaf of vector spaces* on the manifold M .

U_i denotes one of the charts in the atlas for M . The extended atlas defines a topological manifold and the differentiability of the transition maps define a differentiable structure on the tangent bundle manifold.

The tangent bundle is where tangent vectors live, and is itself a smooth manifold. The so-called *Lagrangian* is a natural energy function on the tangent bundle.

Associated with every point x on a smooth manifold M is a *tangent space* $T_x M$ and its *dual*, the *cotangent space* $T_x^* M$. The former consists of the possible directional derivatives, and the latter of the differentials, which can be thought of as infinitesimal elements of the manifold. These spaces always have the same dimension n as the manifold does. The collection of all tangent spaces can in turn be made into a manifold, the tangent bundle, whose dimension is $2n$.

1.1.4.2 Cotangent Bundle of a Smooth Manifold

Recall that the dual of a vector space is the set of linear functionals (i.e., real valued linear functions) on the vector space. In particular, if the vector space is finite and has an inner product then the linear functionals can be realized by the functions $f_v(w) = \langle v, w \rangle$.

The cotangent bundle T^*M is the dual to the tangent bundle TM in the sense that each tangent space has a dual cotangent space as a vector space. The cotangent bundle T^*M is a smooth manifold itself, whose dimension is $2n$. The so-called Hamiltonian is a natural energy function on the cotangent bundle. The total space of a cotangent bundle naturally has the structure of a *symplectic manifold* (see below).

1.1.4.3 Fibre-, Tensor-, and Jet-Bundles

A *fibre bundle* is a space which locally looks like a product of two spaces but may possess a different global structure. Tangent and cotangent bundles are special cases of a fibre bundle.¹⁴

A *tensor bundle* is a direct sum of all tensor products of the tangent bundle and the cotangent bundle.¹⁵ To do calculus on the tensor bundle

¹⁴Every fiber bundle consists of a continuous surjective map: $\pi : E \rightarrow B$, where small regions in the total space E look like small regions in the product space $B \times F$. Here B is called the *base space* while F is the *fiber space*. For example, the *product space* $B \times F$, equipped with π equal to projection onto the first coordinate, is a fiber bundle. This is called the *trivial bundle*. One goal of the theory of bundles is to quantify, via algebraic invariants, what it means for a bundle to be non-trivial.

¹⁵Recall that a *tensor* is a certain kind of geometrical entity which generalizes the

a *connection* is needed (see below). In particular, the *exterior calculus* on a totally antisymmetric tensor bundle allows for a generalization of the classical gradient, divergence and curl operators.

A *jet bundle* is a generalization of both the tangent bundle and the cotangent bundle. The Jet bundle is a certain construction which makes a new smooth fiber bundle out of a given smooth fiber bundle. It makes it possible to write differential equations on sections of a fiber bundle in an invariant form. In contrast with Riemannian manifolds and their (co)tangent bundles, a connection is a *tensor* on the jet bundle.

1.1.5 *Riemannian Manifolds: Configuration Spaces for Lagrangian Mechanics*

To measure distances and angles on manifolds, the manifold must be Riemannian. A *Riemannian manifold* is an analytic manifold in which each *tangent space* is equipped with an *inner product* $g = \langle \cdot, \cdot \rangle$, in a manner

concepts of scalar, vector and linear operator in a way that is independent of any chosen frame of reference. While tensors can be represented by multi-dimensional arrays of components, the point of having a tensor theory is to explain the further implications of saying that a quantity is a tensor, beyond that specifying it requires a number of indexed components. In particular, tensors behave in special ways under coordinate transformations. The tensor notation (also called the covariant formalism) was developed around 1890 by Gregorio Ricci-Curbastro under the title ‘Absolute Differential Geometry’, and made accessible to many mathematicians by the publication of Tullio Levi-Civita’s classic text ‘The Absolute Differential Calculus’ in 1900. The tensor calculus achieved broader acceptance with the introduction of Einstein’s general relativity theory, around 1915. General Relativity is formulated completely in the language of tensors, which Einstein had learned from Levi-Civita himself with great difficulty. But tensors are used also within other fields such as continuum mechanics (e.g., the strain tensor). Note that the word ‘tensor’ is often used as a shorthand for ‘tensor-field’, which is a tensor value defined at every point in a manifold.

The so-called ‘classical approach’ views tensors as multidimensional arrays that are nD generalizations of scalars, 1D vectors and 2D matrices. The ‘components’ of the tensor are the indices of the array. This idea can then be further generalized to tensor-fields, where the elements of the tensor are functions, or even differentials.

On the other hand, the so-called ‘modern’ or component-free approach, views tensors initially as abstract geometrical objects, expressing some definite type of multi-linear concept. Their well-known properties can be derived from their definitions, as linear maps, or more generally; and the rules for manipulations of tensors arise as an extension of linear algebra to multilinear algebra. This treatment has largely replaced the component-based treatment for advanced study, in the way that the more modern component-free treatment of vectors replaces the traditional component-based treatment after the component-based treatment has been used to provide an elementary motivation for the concept of a vector. You could say that the slogan is ‘tensors are elements of some tensor bundle’.

which varies smoothly from point to point. Given two tangent vectors X and Y , the inner product $\langle X, Y \rangle$ gives a real number. The dot (or scalar) product is a typical example of an inner product. This allows one to define various notions such as length, angles, areas (or, volumes), curvature, gradients of functions and divergence of vector-fields. Most familiar curves and surfaces, including n -spheres and Euclidean space, can be given the structure of a Riemannian manifold.

Any smooth manifold admits a Riemannian metric, which often helps to solve problems of differential topology. It also serves as an entry level for the more complicated structure of pseudo-Riemannian manifolds, which (in four dimensions) are the main objects of the general relativity theory.¹⁶

Every smooth submanifold of \mathbb{R}^n (see extrinsic view above) has an induced Riemannian metric g : the inner product on each tangent space is the restriction of the inner product on \mathbb{R}^n . Therefore, one could define a Riemannian manifold as a *metric space* which is *isometric* to a smooth submanifold of \mathbb{R}^n with the induced intrinsic metric, where isometry here is meant in the sense of preserving the length of curves.

Usually a Riemannian manifold M is defined as a smooth manifold with a smooth section of positive-definite quadratic forms on the associated *tangent bundle* TM . Then one has to work to show that it can be turned to a metric space.

Even though Riemannian manifolds are usually ‘curved’ (e.g., the space-time of general relativity), there is still a notion of ‘straight line’ on them: the *geodesics*. These are curves which locally join their points along shortest paths.

In Riemannian manifolds, the notions of geodesic completeness, topological completeness and metric completeness are the same: that each implies the other is the content of the *Hopf-Rinow Theorem*.

1.1.5.1 Riemann Surfaces

A *Riemann surface*, is a 1D complex manifold. Riemann surfaces can be thought of as ‘deformed versions’ of the complex plane: locally near every point they look like patches of the complex plane, but the global topology can be quite different. For example, they can look like a sphere, or a torus, or a couple of sheets glued together.

¹⁶A pseudo-Riemannian manifold is a variant of Riemannian manifold where the metric tensor is allowed to have an indefinite signature (as opposed to a positive-definite one).

The main point of Riemann surfaces is that holomorphic (analytic complex) functions may be defined between them. Riemann surfaces are nowadays considered the natural setting for studying the global behavior of these functions, especially multi-valued functions such as the square root or the logarithm.

Every Riemann surface is a 2D real analytic manifold (i.e., a surface), but it contains more structure (specifically, a complex structure) which is needed for the unambiguous definition of holomorphic functions. A 2D real manifold can be turned into a Riemann surface (usually in several inequivalent ways) iff it is *orientable*. So the sphere and torus admit complex structures, but the Möbius strip, Klein bottle and projective plane do not.

Geometrical facts about Riemann surfaces are as ‘nice’ as possible, and they often provide the intuition and motivation for generalizations to other curves and manifolds. The Riemann–Roch Theorem is a prime example of this influence.¹⁷

Examples of Riemann surfaces include: the complex plane¹⁸, open subsets of the complex plane¹⁹, *Riemann sphere*²⁰, and many others.

Riemann surfaces naturally arise in string theory as models of string

¹⁷Formally, let X be a Hausdorff space. A homeomorphism from an open subset $U \subset X$ to a subset of \mathbb{C} is a chart. Two charts f and g whose domains intersect are said to be compatible if the maps $f \circ g^{-1}$ and $g \circ f^{-1}$ are holomorphic over their domains. If A is a collection of compatible charts and if any $x \in X$ is in the domain of some $f \in A$, then we say that A is an atlas. When we endow X with an atlas A , we say that (X, A) is a Riemann surface.

Different atlases can give rise to essentially the same Riemann surface structure on X ; to avoid this ambiguity, one sometimes demands that the given atlas on X be maximal, in the sense that it is not contained in any other atlas. Every atlas A is contained in a unique maximal one by *Zorn’s lemma*.

¹⁸The complex plane \mathbb{C} is perhaps the most trivial Riemann surface. The map $f(z) = z$ (the identity map) defines a chart for \mathbb{C} , and f is an atlas for \mathbb{C} . The map $g(z) = z^*$ (the conjugate map) also defines a chart on \mathbb{C} and g is an atlas for \mathbb{C} . The charts f and g are not compatible, so this endows \mathbb{C} with two distinct Riemann surface structures.

¹⁹In a fashion analogous to the complex plane, every open subset of the complex plane can be viewed as a Riemann surface in a natural way. More generally, every open subset of a Riemann surface is a Riemann surface.

²⁰The Riemann sphere is a useful visualization of the extended complex plane, which is the complex plane plus a point at infinity. It is obtained by imagining that all the rays emanating from the origin of the complex plane eventually meet again at a point called the point at infinity, in the same way that all the meridians from the south pole of a sphere get to meet each other at the north pole.

Formally, the Riemann sphere is obtained via a one-point compactification of the complex plane. This gives it the topology of a 2-sphere. The sphere admits a unique complex structure turning it into a Riemann surface. The Riemann sphere can be characterized as the unique simply-connected, compact Riemann surface.

interactions.

1.1.5.2 Riemannian Geometry

Riemannian geometry is the study of smooth manifolds with Riemannian metrics g , i.e. a choice of positive-definite quadratic form $g = \langle \cdot, \cdot \rangle$ on a manifold's tangent spaces which varies smoothly from point to point. This gives in particular local ideas of angle, length of curves, and volume. From those some other global quantities can be derived by integrating local contributions.

The manifold may also be given an *affine connection*,²¹ which is roughly an idea of change from one point to another. If the metric does not 'vary from point to point' under this connection, we say that the metric and connection are compatible, and we have a Riemann–Cartan manifold. If this connection is also self-commuting when acting on a scalar function, we say that it is torsion-free, and the manifold is a Riemannian manifold.

The Levi–Civita Connection

In Riemannian geometry, the *Levi–Civita connection* (named after Tullio Levi–Civita) is the torsion-free Riemannian connection, i.e., a torsion-free connection of the tangent bundle, preserving a given Riemannian metric (or, pseudo-Riemannian metric).²² The fundamental Theorem of Riemann–

²¹Connection (or, covariant derivative) is a way of specifying a derivative of a vector-field along another vector-field on a manifold. That is an application to tangent bundles; there are more general connections, used to formulate intrinsic differential equations. Connections give rise to parallel transport along a curve on a manifold. A connection also leads to invariants of curvature, and the so-called *torsion*.

An affine connection is a connection on the tangent bundle TM of a smooth manifold M . In general, it might have a non-vanishing torsion.

The curvature of a connected manifold can be characterized intrinsically by taking a vector at some point and *parallel transporting* it along a curve on the manifold. Although comparing vectors at different points is generally not a well-defined process, an affine connection ∇ is a rule which describes how to legitimately move a vector along a curve on the manifold without changing its direction ('keeping the vector parallel').

²²Formally, let (M, g) be a Riemannian manifold (or pseudo-Riemannian manifold); then an affine connection is the Levi–Civita connection if it satisfies the following conditions:

- (1) *Preserves metric* g , i.e., for any three vector-fields $X, Y, Z \in M$ we have $Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z)$, where $Xg(Y, Z)$ denotes the derivative of a function $g(Y, Z)$ along a vector-field X .
- (2) *Torsion-free*, i.e., for any two vector-fields $X, Y, Z \in M$ we have $\nabla_X Y - \nabla_Y X = [X, Y]$, where $[X, Y]$ is the *Lie bracket* for vector-fields X and Y .

nian geometry states that there is a unique connection which satisfies these properties.²³

In the theory of Riemannian and pseudo-Riemannian manifolds the term covariant derivative is often used for the Levi-Civita connection. The coordinate expression of the connection is given by *Christoffel symbols*.²⁴ Note that connection is not a tensor, except on jet bundles.

The Fundamental Riemannian Tensors

The two basic objects in Riemannian geometry are the metric tensor and the curvature tensor. The *metric tensor* $g = \langle \cdot, \cdot \rangle$ is a symmetric second-order (i.e., $(0, 2)$) tensor that is used to measure distance in a space. In other words, given a Riemannian manifold, we make a choice of a $(0, 2)$ -tensor on the manifold's tangent spaces.²⁵ At a given point in the manifold, this tensor takes a pair of vectors in the tangent space to that point, and gives a

²³The Levi-Civita connection defines also a derivative along curves, usually denoted by D . Given a smooth curve (a path) $\gamma = \gamma(t) : \mathbb{R} \rightarrow M$ and a vector-field $X = X(t)$ on γ , its derivative along γ is defined by: $D_t X = \nabla_{\dot{\gamma}(t)} X$. This equation defines the *parallel transport* for a vector-field X .

²⁴The Christoffel symbols, named for Elwin Bruno Christoffel (1829–1900), are coordinate expressions for the Levi-Civita connection derived from the metric tensor. The Christoffel symbols are used whenever practical calculations involving geometry must be performed, as they allow very complex calculations to be performed without confusion. In particular, if we denote the unit vectors on M as $e_i = \partial/\partial x_i$, then the Christoffel symbols of the second kind are defined by $\Gamma_{ij}^k = \langle \nabla_{e_i} e_j, e_k \rangle$. Alternatively, using the metric tensor g_{ik} (see below) we get the explicit expression for the Christoffel symbols in a holonomic coordinate basis:

$$\Gamma_{kl}^i = \frac{1}{2} g^{im} \left(\frac{\partial g_{mk}}{\partial x^l} + \frac{\partial g_{ml}}{\partial x^k} - \frac{\partial g_{kl}}{\partial x^m} \right).$$

In a general, nonholonomic coordinates they include the additional commutation coefficients. The Christoffel symbols are used to define the covariant derivative of various tensor-fields, as well as the Riemannian curvature. Also, they figure in the *geodesic equation*:

$$\frac{d^2 x^i}{dt^2} + \Gamma_{jk}^i \frac{dx^j}{dt} \frac{dx^k}{dt} = 0$$

for the curve $x^i = x^i(t)$ on the smooth manifold M .

²⁵The most familiar example is that of basic high-school geometry: the 2D Euclidean metric tensor, in the usual $x - y$ coordinates, reads: $g = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. The associated length of a curve is given by the familiar calculus formula: $L = \int_a^b \sqrt{(dx)^2 + (dy)^2}$.

The unit sphere in \mathbb{R}^3 comes equipped with a natural metric induced from the ambient Euclidean metric. In standard spherical coordinates (θ, ϕ) the metric takes the form:

$$g = \begin{bmatrix} 1 & 0 \\ 0 & \sin^2 \theta \end{bmatrix}, \text{ which is usually written as: } g = d\theta^2 + \sin^2 \theta d\phi^2.$$

real number. This concept is just like a dot product, or inner product. This function from vectors into the real numbers is required to vary smoothly from point to point.²⁶

On any Riemannian manifold, from its second-order metric tensor $g = \langle \cdot, \cdot \rangle$, one can derive the associated fourth-order *Riemann curvature tensor*. This tensor is the most standard way to express curvature of Riemannian manifolds, or more generally, any manifold with an affine connection, torsionless or with torsion.²⁷

²⁶Once a local coordinate system x^i is chosen, the metric tensor appears as a matrix, conventionally given by its components, $g = g_{ij}$. Given the metric tensor of a Riemannian manifold and using the Einstein summation notation for implicit sums, the length of a segment of a curve parameterized by t , from a to b , is defined as:

$L = \int_a^b \sqrt{g_{ij} \frac{dx^i}{dt} \frac{dx^j}{dt}} dt$. Also, the angle θ between two tangent vectors u^i, v^i is defined as: $\cos \theta = \frac{g_{ij} u^i v^j}{\sqrt{|g_{ij} u^i u^j| |g_{ij} v^i v^j|}}$.

²⁷The Riemann curvature tensor is given in terms of a Levi-Civita connection ∇ (more generally, an affine connection, or covariant differentiation, see below) by the following formula:

$$R(u, v)w = \nabla_u \nabla_v w - \nabla_v \nabla_u w - \nabla_{[u, v]} w,$$

where u, v, w are tangent vector-fields and $R(u, v)$ is a linear transformation of the tangent space of the manifold; it is linear in each argument. If $u = \partial/\partial x_i$ and $v = \partial/\partial x_j$ are coordinate vector-fields then $[u, v] = 0$ and therefore the above formula simplifies to

$$R(u, v)w = \nabla_u \nabla_v w - \nabla_v \nabla_u w,$$

i.e., the curvature tensor measures non-commutativity of the covariant derivative. The linear transformation $w \mapsto R(u, v)w$ is also called the curvature transformation or endomorphism.

In local coordinates x^μ (e.g., in general relativity) the Riemann curvature tensor can be written using the Christoffel symbols of the manifold's Levi-Civita connection:

$$R^\rho{}_{\sigma\mu\nu} = \partial_\mu \Gamma^\rho_{\nu\sigma} - \partial_\nu \Gamma^\rho_{\mu\sigma} + \Gamma^\rho_{\mu\lambda} \Gamma^\lambda_{\nu\sigma} - \Gamma^\rho_{\nu\lambda} \Gamma^\lambda_{\mu\sigma}.$$

The Riemann curvature tensor has the following symmetries:

$$\begin{aligned} R(u, v) &= -R(v, u), & \langle R(u, v)w, z \rangle &= -\langle R(u, v)z, w \rangle, \\ R(u, v)w + R(v, w)u + R(w, u)v &= 0. \end{aligned}$$

The last identity was discovered by Ricci, but is often called the first Bianchi identity or algebraic Bianchi identity, because it looks similar to the Bianchi identity below. These three identities form a complete list of symmetries of the curvature tensor, i.e. given any tensor which satisfies the identities above, one can find a Riemannian manifold with such a curvature tensor at some point. Simple calculations show that such a tensor has $n^2(n^2 - 1)/12$ independent components.

The *Bianchi identity* involves the covariant derivatives:

$$\nabla_u R(v, w) + \nabla_v R(w, u) + \nabla_w R(u, v) = 0.$$

1.1.5.3 APPLICATION: *Lagrangian Mechanics*

Riemannian manifolds are a natural stage for the *Lagrangian mechanics*, which is a re-formulation of classical mechanics introduced by Joseph Louis Lagrange in 1788. In Lagrangian mechanics, the trajectory of an object is derived by finding the path which minimizes the action, a quantity which is the integral of the Lagrangian over time. The Lagrangian for classical mechanics L is taken to be the difference between the kinetic energy T and the potential energy V , so $L = T - V$. This considerably simplifies many physical problems.

For example, consider a bead on a hoop. If one were to calculate the motion of the bead using *Newtonian mechanics*, one would have a complicated set of equations which would take into account the forces that the hoop exerts on the bead at each moment. The same problem using Lagrangian mechanics is much simpler. One looks at all the possible motions that the bead could take on the hoop and mathematically finds the one which *minimizes the action*. There are fewer equations since one is not directly calculating the influence of the hoop on the bead at a given moment.

Lagrange's Equations

The equations of motion in Lagrangian mechanics are *Lagrange's equations*, also known as *Euler-Lagrange equations*. Below, we sketch out the derivation of Lagrange's equation from Newton's laws of motion (see next chapter for details).

Consider a single mechanical particle with mass m and position vector \vec{r} . The applied force, \vec{F} , can be expressed as the gradient (denoted ∇) of a scalar potential energy function $V(\vec{r}, t)$:

$$\vec{F} = -\nabla V.$$

A contracted curvature tensor is called the *Ricci tensor*. It is a symmetric second-order tensor given by:

$$R_{ik} = \frac{\partial \Gamma^l_{ik}}{\partial x^l} - \frac{\partial \Gamma^l_{il}}{\partial x^k} + \Gamma^l_{ik} \Gamma^m_{lm} - \Gamma^m_{il} \Gamma^l_{km}.$$

Its further contraction gives the *Ricci scalar curvature*, $R = g^{ik} R_{ik}$. The *Einstein tensor* G_{ik} is defined in terms of the Ricci tensor R_{ik} and the Ricci scalar R ,

$$G_{ik} = R_{ik} - \frac{1}{2} g_{ik} R.$$

Such a force is independent of third- or higher-order derivatives of \vec{r} , so *Newton's Second Law* forms a set of 3 second-order ODEs. Therefore, the motion of the particle can be completely described by 6 independent variables, or degrees of freedom (DOF). An obvious set of variables is the Cartesian components of \vec{r} and their time derivatives, at a given instant of time, that is position (x, y, z) and velocity (v_x, v_y, v_z) .

More generally, we can work with a set of generalized coordinates, q^i , ($i = 1, \dots, n$), and their time derivatives, the generalized velocities, \dot{q}^i . The position vector \vec{r} is related to the generalized coordinates by some transformation equation: $\vec{r} = \vec{r}(q^i, t)$. The term ‘generalized coordinates’ is really a leftover from the period when Cartesian coordinates were the default coordinate system. In the q^i -coordinates the Lagrange’s equations read:

$$\frac{\partial L}{\partial q^i} = \frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i},$$

where $L = T - V$ is the system’s Lagrangian.

The time integral of the Lagrangian L , denoted S is called the *action*:²⁸

$$S = \int L dt.$$

Let q_0 and q_1 be the coordinates at respective initial and final times t_0 and t_1 . Using the *calculus of variations*, it can be shown the Lagrange’s equations are equivalent to the *Hamilton’s principle*: “The system undergoes the trajectory between t_0 and t_1 whose action has a stationary value.” This is formally written:

$$\delta S = 0,$$

where by ‘stationary’, we mean that the action does not vary to first-order for infinitesimal deformations of the trajectory, with the end-points (q_0, t_0)

²⁸The action principle is an assertion about the nature of motion, from which the trajectory of a dynamical system subject to some forces can be determined. The path of an object is the one that yields a stationary value for a quantity called the action. Thus, instead of thinking about an object accelerating in response to applied forces, one might think of them picking out the path with a stationary action. The action is a scalar (a number) with the unit of measure for Action as Energy \times Time. Although equivalent in classical mechanics with Newton’s laws, the action principle is better suited for generalizations and plays an important role in modern physics. Indeed, this principle is one of the great generalizations in physical science. In particular, it is fully appreciated and best understood within quantum mechanics. Richard Feynman’s path integral formulation of quantum mechanics is based on a stationary-action principle, using path integrals. Maxwell’s equations can be derived as conditions of stationary action.

and (q_1, t_1) fixed.²⁹

The total energy function called *Hamiltonian*, denoted by H , is obtained by performing a *Legendre transformation* on the Lagrangian.³⁰ The Hamiltonian is the basis for an alternative formulation of classical mechanics known as Hamiltonian mechanics (see below).

In 1948, R.P. Feynman invented the *path-integral formulation* extending the principle of least action to quantum mechanics for electrons and pho-

²⁹More generally, a Lagrangian $\mathcal{L}[\varphi^i]$ of a dynamical system is a function of the dynamical variables $\varphi^i(x)$ and concisely describes the equations of motion of the system in coordinates x^i , ($i = 1, \dots, n$). The equations of motion are obtained by means of an *action principle*, written as

$$\frac{\delta S}{\delta \varphi_i} = 0,$$

where the *action* is a functional

$$S[\varphi^i] = \int \mathcal{L}[\varphi^i(s)] d^n x,$$

($d^n x = dx^1 \dots dx^n$).

The equations of motion obtained by means of the functional derivative are identical to the usual Euler–Lagrange equations. Dynamical system whose equations of motion are obtainable by means of an action principle on a suitably chosen Lagrangian are known as Lagrangian dynamical systems. Examples of Lagrangian dynamical systems range from the (classical version of the) Standard Model, to Newton’s equations, to purely mathematical problems such as geodesic equations and the Plateau’s problem.

The Lagrangian mechanics is important not just for its broad applications, but also for its role in advancing deep understanding of physics. Although Lagrange sought to describe classical mechanics, the action principle that is used to derive the Lagrange’s equation is now recognized to be deeply tied to quantum mechanics: physical action and quantum–mechanical phase (waves) are related via Planck’s constant, and the *Principle of stationary action* can be understood in terms of constructive interference of wave functions. The same principle, and the Lagrangian formalism, are tied closely to *Noether Theorem*, which relates physical conserved quantities to continuous symmetries of a physical system; and Lagrangian mechanics and Noether’s Theorem together yield a natural formalism for first quantization by including commutators between certain terms of the Lagrange’s equations of motion for a physical system.

More specifically, in field theory, occasionally a distinction is made between the Lagrangian L , of which the action is the time integral $S = \int L dt$ and the Lagrangian density \mathcal{L} , which one integrates over all space–time to get the 4D action:

$$S[\varphi^i] = \int \mathcal{L}[\varphi^i(x)] d^4 x.$$

The Lagrangian is then the spatial integral of the Lagrangian density.

³⁰The Hamiltonian is the Legendre transform of the Lagrangian:

$$H(q, p, t) = \sum_i \dot{q}^i p_i - L(q, \dot{q}, t).$$

tons. In this formulation, particles travel every possible path between the initial and final states; the probability of a specific final state is obtained by summing over all possible trajectories leading to it. In the classical regime, the path integral formulation cleanly reproduces the Hamilton's principle, as well as the Fermat's principle in optics.

1.1.5.4 Finsler manifolds

Finsler manifolds represent generalization of Riemannian manifolds. A Finsler manifold allows the definition of distance, but not of angle; it is an analytic manifold in which each tangent space is equipped with a norm $\|\cdot\|$ in a manner which varies smoothly from point to point. This norm can be extended to a metric, defining the length of a curve; but it cannot in general be used to define an inner product. Any Riemannian manifold (but not a pseudo-Riemannian manifold) is a Finsler manifold.³¹

1.1.6 Symplectic Manifolds: Phase-Spaces for Hamiltonian Mechanics

A *symplectic manifold* is a smooth manifold M equipped with a closed, non-degenerate, 2-form ω called the *symplectic volume form*, or *Liouville measure*. This condition forces symplectic manifolds to be even-dimensional. Cotangent bundles, which arise as phase-spaces in *Hamiltonian mechanics*, are the motivating example, but many compact manifolds also have symplectic structure. All surfaces have a symplectic structure, since a symplectic structure is simply a volume form. The study of symplectic manifolds is called symplectic geometry/topology.

Symplectic manifolds arise naturally in abstract formulations of classical mechanics as the *cotangent bundles* of configuration manifolds: the set of all possible configurations of a system is modelled as a manifold M , and this manifold's cotangent bundle T^*M describes the *phase-space* of the

³¹Formally, a Finsler manifold is a differentiable manifold M with a Banach norm defined over each tangent space such that the Banach norm as a function of position is smooth, usually it is assumed to satisfy the following regularity condition:

For each point x of M , and for every nonzero vector X in the tangent space $T \times M$, the second derivative of the function $L : T \times M \rightarrow R$ given by $L(w) = \frac{1}{2}\|w\|^2$ at X is positive definite.

The length of a smooth curve γ in a Finsler manifold M is given by $\int \left\| \frac{d\gamma}{dt}(t) \right\| dt$.

Length is invariant under reparametrization. With the above regularity condition, *geodesics* are locally length-minimizing curves with constant speed, or equivalently curves in whose *energy function*, $\int \left\| \frac{d\gamma}{dt}(t) \right\|^2 dt$, is *extremal* under functional derivatives.

dynamical system.³²

Any real-valued differentiable function H on a symplectic manifold can serve as an *energy function* or *Hamiltonian*.³³ Associated to any Hamiltonian is a *Hamiltonian vector-field*.³⁴ The integral curves of the Hamiltonian

³²*Hamiltonian mechanics* is a re-formulation of classical mechanics that was invented in 1833 by William Rowan Hamilton. It arose from *Lagrangian mechanics*, a previous re-formulation of classical Newtonian mechanics, introduced by J.L. Lagrange in 1788. It can however be formulated without recourse to Lagrangian mechanics, using symplectic manifolds. It is based on *canonical Hamilton's equations of motion*:

$$\dot{p} = -\frac{\partial H}{\partial q}, \quad \dot{q} = \frac{\partial H}{\partial p},$$

with canonical coordinate q and momentum p variables and the Hamiltonian (total energy) function $H = H(q, p)$. An important special case consists of those Hamiltonians that are quadratic forms, that is, Hamiltonians that can be written as:

$$H(q, p) = \frac{1}{2} \langle p, p \rangle_q$$

This Hamiltonian consists entirely of the kinetic term. If one considers a Riemannian manifold, or a pseudo-Riemannian manifold, so that one has an invertible, non-degenerate metric, then $\langle \cdot, \cdot \rangle_q$ is simply the inverse of the Riemannian metric g .

³³In contrast, the quantum-mechanical Hamiltonian H is the observable corresponding to the total energy of the system. It generates the time evolution of quantum states (see, e.g., [Ivancevic and Ivancevic (2006)]). If the wave-function $|\psi(t)\rangle$ represents the state of the quantum system at time t , then its time evolution is given by the Schrödinger equation:

$$H |\psi(t)\rangle = i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle,$$

where \hbar is the Planck constant. If the quantum Hamiltonian H is independent of time, then the time evolution is given by:

$$|\psi(t)\rangle = \exp\left(-\frac{iHt}{\hbar}\right) |\psi(0)\rangle.$$

³⁴In canonical coordinates (q^i, p_i) on a symplectic manifold M , the symplectic form ω can be written as $\omega = \sum_n dq^i \wedge dp_i$ and thus the Hamiltonian vector-field takes the form $X_H = \left(\frac{\partial H}{\partial p_i}, -\frac{\partial H}{\partial q^i}\right)$. The Hamiltonian vector-field X_H also induces a special operation, the *Poisson bracket*, which is a bilinear map turning two differentiable functions on a symplectic manifold M into a function on M . In particular, if we have two functions, f and g , then the Poisson bracket $\{f, g\} = \omega(df, dg)$. In canonical coordinates (q^i, p_i) on the phase-space manifold, the Poisson bracket takes the form

$$\{f, g\} = \omega(df, dg) = \sum_{i=1}^n \left[\frac{\partial f}{\partial q^i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q^i} \right].$$

The time evolution of a function f on the symplectic manifold can be given as a one-parameter family of symplectomorphisms, with the time t being the parameter. The total time derivative can be written as $\frac{d}{dt} f = \frac{\partial}{\partial t} f + \{f, H\} = \frac{\partial}{\partial t} f - \{H, f\}$. The Poisson

vector-field are solutions to the *Hamilton–Jacobi equation*.³⁵

The Hamiltonian vector-field defines a flow on the symplectic manifold, called a *Hamiltonian flow* or *symplectomorphism*. By *Liouville’s Theorem*, Hamiltonian flows preserve the volume form on the phase space.³⁶

Symplectic manifolds are special cases of a Poisson manifold; the definition of a symplectic manifold requires that the symplectic 2-form ω be non-degenerate everywhere. If this condition is violated, the manifold may still be a *Poisson manifold*.³⁷ Also, a symplectic manifold endowed with a metric that is compatible with the symplectic form is a *Kähler manifold*.

bracket acts on functions on the symplectic manifold, thus giving the space of functions on the manifold the structure of a *Lie algebra* (see Lie groups below).

³⁵The Hamilton–Jacobi equation (HJE) is a particular canonical transformation of the classical Hamiltonian which results in a first order, nonlinear differential equation whose solution describes the behavior of the system. While the canonical Hamilton’s equations of motion represent the system of first order ODEs, two for each coordinate, the HJE is a single PDE of one variable for each coordinate. If we have a Hamiltonian of the form then the HJE for that system is

$$H\left(q^1, \dots, q^n; \frac{\partial S}{\partial q^1}, \dots, \frac{\partial S}{\partial q^n}; t\right) + \frac{\partial S}{\partial t} = 0,$$

where S represents the classical action functional. The HJE can be used to solve several problems elegantly, such as the *Kepler problem*.

³⁶Liouville’s Theorem, named after the French mathematician Joseph Liouville, is a key Theorem in classical statistical and Hamiltonian mechanics. It asserts that the phase-space distribution function is constant along the trajectories of the system – that is that the density of system points in the vicinity of a given system point travelling through phase-space is constant with time.

The *Liouville equation* describes the time evolution of a phase-space distribution function, or Liouville measure. Consider a dynamical system with coordinates q^i and conjugate momenta p_i ($i = 1, \dots, n$). The time evolution of the phase-space distribution $\rho(p, q)$ is governed by the Liouville equation:

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \sum_{i=1}^n \left(\frac{\partial \rho}{\partial q^i} \dot{q}^i + \frac{\partial \rho}{\partial p_i} \dot{p}_i \right) = 0.$$

Time derivatives are denoted by dots, and are evaluated according to *Hamilton’s equations* for the system. The Liouville’s Theorem states that the phase-space distribution function $\rho(p, q)$ is constant along any trajectory in phase space. The Theorem is often restated in terms of the *Poisson bracket* with the Hamiltonian function $H = H(q, p)$:

$$\frac{\partial \rho}{\partial t} = -\{\rho, H\}.$$

Geometrically, this Theorem says that Liouville measure is invariant under the *Hamiltonian flow*.

³⁷Closely related to even-dimensional symplectic manifolds are the odd-dimensional manifolds known as *contact manifolds*. Any $(2n + 1)$ -D contact manifold (M, ω) gives rise to a $(2n + 2)$ -D symplectic manifold $(M \times \mathbb{R}, d(e^t \omega))$.

There are several natural geometric notions of *submanifold* of a symplectic manifold. There are symplectic submanifolds (potentially of any even dimension), where the symplectic form is required to induce a symplectic form on the submanifold. The most important case of these is that of *Lagrangian submanifold*, which are isotropic submanifolds of maximal dimension, namely half the dimension of the ambient manifold. Lagrangian submanifolds arise naturally in many physical and geometric situations.³⁸

1.1.7 Lie Groups

A Lie group is smooth manifold which also carries a group structure whose product and inversion operations are smooth as maps of manifolds. These objects arise naturally in describing symmetries.

A Lie group is a group whose elements can be continuously parametrized by real numbers, such as the *rotation group* $SO(3)$, which can be parametrized by the *Euler angles*. More formally, a Lie group is an analytic real or complex manifold that is also a group, such that the group operations multiplication and inversion are analytic maps. Lie groups are important in mathematical analysis, physics and geometry because they serve to describe the symmetry of analytical structures. They were introduced by *Sophus Lie* in 1870 in order to study symmetries of differential equations.

While the Euclidean space \mathbb{R}^n is a *real Lie group* (with ordinary vector addition as the group operation), more typical examples are given by matrix Lie groups, i.e., groups of invertible matrices (under matrix multiplication). For instance, the group $SO(3)$ of all rotations in \mathbb{R}^3 is a matrix Lie group.

One classifies Lie groups regarding their *algebraic properties*³⁹ (simple, semisimple, solvable, nilpotent, Abelian), their *connectedness* (connected

³⁸One major example is that the graph of a symplectomorphism in the product symplectic manifold $(M \times M, \omega \times -\omega)$ is Lagrangian. Their intersections display rigidity properties not possessed by smooth manifolds; the *Arnold conjecture* gives the sum of the submanifold's *Betti numbers* as a lower bound for the number of self intersections of a smooth Lagrangian submanifold, rather than the *Euler characteristic* in the smooth case.

³⁹If G and H are Lie groups (both real or both complex), then a *Lie-group-homomorphism* $f : G \rightarrow H$ is a *group homomorphism* which is also an *analytic map* (one can show that it is equivalent to require only that f be continuous). The composition of two such homomorphisms is again a homomorphism, and the class of all (real or complex) Lie groups, together with these morphisms, forms a *category*. The two Lie groups are called *isomorphic* iff there exists a bijective homomorphism between them whose inverse is also a homomorphism. Isomorphic Lie groups do not need to be distinguished for all practical purposes; they only differ in the notation of their elements.

or simply connected) and their *compactness*.⁴⁰

To every Lie group, we can associate a *Lie algebra* which completely captures the local structure of the group (at least if the Lie group is connected).⁴¹

⁴⁰An n -torus $T^n = S^1 \times S^1 \times \cdots \times S^1$ (as defined above) is an example of a *compact Abelian Lie group*. This follows from the fact that the unit circle S^1 is a compact Abelian Lie group (when identified with the unit complex numbers with multiplication). Group multiplication on T^n is then defined by coordinate-wise multiplication.

Toroidal groups play an important part in the theory of compact Lie groups. This is due in part to the fact that in any compact Lie group one can always find a maximal torus; that is, a closed subgroup which is a torus of the largest possible dimension.

⁴¹Conventionally, one can regard any field X of tangent vectors on a Lie group as a partial differential operator, denoting by Xf the *Lie derivative* (the *directional derivative*) of the scalar field f in the direction of X . Then a vector-field on a Lie group G is said to be left-invariant if it commutes with left translation, which means the following. Define $L_g[f](x) = f(gx)$ for any analytic function $f : G \rightarrow \mathbb{R}$ and all $g, x \in G$. Then the vector-field X is left-invariant iff $XL_g = L_gX$ for all $g \in G$. Similarly, instead of \mathbb{R} , we can use \mathbb{C} . The set of all vector-fields on an analytic manifold is a *Lie algebra* over \mathbb{R} (or \mathbb{C}).

On a Lie group G , the left-invariant vector-fields form a subalgebra, the Lie algebra \mathfrak{g} associated with G . This Lie algebra is finite-dimensional (it has the same dimension as the manifold G) which makes it susceptible to classification attempts. By classifying \mathfrak{g} , one can also get a handle on the group G . The representation theory of simple Lie groups is the best and most important example.

Every element v of the tangent space T_e at the identity element e of G determines a unique left-invariant vector-field whose value at the element g of G is denoted by gv ; the vector space underlying the Lie algebra \mathfrak{g} may therefore be identified with T_e .

Every vector-field v in the Lie algebra \mathfrak{g} determines a function $c : \mathbb{R} \rightarrow G$ whose derivative everywhere is given by the corresponding left-invariant vector-field: $c'(t) = TL_{c(t)}v$ and which has the property: $c(s+t) = c(s)c(t)$, (for all s and t) (the operation on the r.h.s. is the group multiplication in G). The formal similarity of this formula with the one valid for the elementary exponential function justifies the definition: $\exp(v) = c(1)$. This is called the *exponential map*, and it maps the Lie algebra \mathfrak{g} into the Lie group G . It provides a *diffeomorphism* between a neighborhood of 0 in \mathfrak{g} and a neighborhood of e in G . This exponential map is a generalization of the exponential function for real numbers (since \mathbb{R} is the Lie algebra of the Lie group of positive real numbers with multiplication), for complex numbers (since \mathbb{C} is the Lie algebra of the Lie group of non-zero complex numbers with multiplication) and for matrices (since $M(n, \mathbb{R})$ with the regular commutator is the Lie algebra of the Lie group $GL(n, \mathbb{R})$ of all invertible matrices). As the exponential map is surjective on some neighborhood N of e , it is common to call elements of the Lie algebra *infinitesimal generators* of the group G .

The exponential map and the Lie algebra determine the local group structure of every connected Lie group, because of the *Baker–Campbell–Hausdorff formula*: there exists a neighborhood U of the zero element of the Lie algebra \mathfrak{g} , such that for $u, v \in U$ we have

$$\exp(u)\exp(v) = \exp(u + v + 1/2[u, v] + 1/12[[u, v], v] - 1/12[[u, v], u] - \dots),$$

where the omitted terms are known and involve *Lie brackets* of four or more elements.

1.1.7.1 APPLICATION: *Physical Examples of Lie Groups*

Here are a few examples of Lie groups and their relations to other areas of mathematics and physics:

- (1) Euclidean space \mathbb{R}^n is an Abelian Lie group (with ordinary vector addition as the group operation).
- (2) The group $GL_n(\mathbb{R})$ of invertible matrices (under matrix multiplication) is a Lie group of dimension n^2 . It has a subgroup $SL_n(\mathbb{R})$ of matrices of determinant 1 which is also a Lie group.
- (3) The group $O_n(\mathbb{R})$ generated by all rotations and reflections of an n D vector space is a Lie group called the *orthogonal group*. It has a subgroup of elements of determinant 1, called the special orthogonal group $SO(n)$, which is the *group of rotations* in \mathbb{R}^n .⁴²
- (4) Spin groups are double covers of the special orthogonal groups (used e.g., for studying fermions in quantum field theory).
- (5) The group $Sp_{2n}(\mathbb{R})$ of all matrices preserving a symplectic form is a Lie group called the *symplectic group*.
- (6) The Lorentz group and the Poincaré group of isometries of space–time are Lie groups of dimensions 6 and 10 that are used in special relativity.
- (7) The Heisenberg group is a Lie group of dimension 3, used in quantum mechanics.
- (8) The unitary group $U(n)$ is a compact group of dimension n^2 consisting of unitary matrices. It has a subgroup of elements of determinant 1, called the special unitary group $SU(n)$.
- (9) The group $U(1) \times SU(2) \times SU(3)$ is a Lie group of dimension $1 + 3 + 8 = 12$ that is the *gauge group* of the *Standard Model* of elementary particles, whose dimension corresponds to: 1 photon + 3 vector bosons + 8 gluons.

In case u and v commute, this formula reduces to the familiar *exponential law*:

$$\exp(u)\exp(v) = \exp(u + v).$$

Every homomorphism $f : G \rightarrow H$ of Lie groups induces a homomorphism between the corresponding Lie algebras \mathfrak{g} and \mathfrak{h} . The association $G \Rightarrow \mathfrak{g}$ is called the *Lie Functor*.

⁴²For example, in terms of orthogonal matrices, the rotations about the standard Cartesian coordinate axes (x, y, z) in \mathbb{R}^3 through an angle ϕ are given by:

$$R_x(\phi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & -\sin \phi \\ 0 & \sin \phi & \cos \phi \end{pmatrix}, \quad R_y(\phi) = \begin{pmatrix} \cos \phi & 0 & \sin \phi \\ 0 & 1 & 0 \\ -\sin \phi & 0 & \cos \phi \end{pmatrix}, \quad R_z(\phi) = \begin{pmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

1.1.8 Application: A Bird View on Modern Physics

1.1.8.1 Three Pillars of 20th Century Physics

In this subsection we make small digression into the field of modern physics, which is the major customer for machinery of differential geometry. Arguably, the three most influential geniuses that shaped the world of the 20th Century physics, and at the same time showed the pathway to the current unified physical ‘theory of everything’, have been:

- (1) In the first third of the Century, it had been *Albert Einstein*.
- (2) In the second third of the Century, it was *Richard Feynman*.
- (3) At the end of the Century – and still today, it has been *Edward Witten*.

It is well-known that *Einstein* had three periods of his scientific career:

- (1) Before 1905, when he formulated Special Relativity in a quick series of papers published in *Annalen der Physik* (the most prestigious physics journal of the time). This early period was dominated by his ‘thought experiments’, i.e., ‘concrete physical images’, described in the language of non-professional mathematics. You can say, it was almost pure visualization. This quick and powerful series of ground-braking papers (with just enough maths to be accepted by scientific community) gave him a reputation of the leading physicist and scientist.⁴³
- (2) Although an original and brilliant theory, Special Relativity was not complete, which was obvious to Einstein. So, he embarked onto the general relativity voyage, incorporating gravitation. Now, for this goal, his maths was not strong enough. He spent 10 years fighting with gravity, using the ‘hard’ Riemannian geometry, and talking to the leading mathematician of the time, *David Hilbert*. At the end, they both submitted the same gravitational equations of general relativity (only derived in different ways) to *Annalen der Physik* in November of 1915.
- (3) Although even today considered as the most elegant physical theory, General Relativity is still not complete: it cannot live together in the same world with quantum mechanics. So, Einstein embarked onto the last journey of his life, the search for *unified field theory* – and he ‘failed’⁴⁴ after 30 years of unsuccessful struggle with a task to big for

⁴³Recall that the Nobel Prize was ‘in the air’ for Einstein for more than 15 years; at the end he got it in 1921, for his discovery of the Photo-Electric Effect.

⁴⁴Einstein ‘failed’ in the same way as Hilbert ‘failed’ with his Program of axiomatic formalization of all mathematical sciences. Their apparent ‘failure’ still influences development of physics and mathematics, apparently converging into superstring theory.

one man.

Feynman's story is very different. All his life he was a profoundly original scientist, similar to the young Einstein. He refused to take anybody's word for anything, which meant that he had to reinvent for himself almost the whole of physics. It took him five years of concentrated work to reinvent quantum mechanics. At the end, he got a new version of quantum mechanics that he (and only he) could understand. In orthodox physics it was said: Suppose an electron is in this state at a certain time, then you calculate its future behavior by solving Schrodinger equation. Instead of this, Feynman said simply: "The electron does whatever it likes." A history of the electron is any possible path in space and time. The behavior of the electron is just the result of adding together all histories according to some simple rules that Feynman worked out. His path-integral and related Feynman diagrams, for long defied rigorous mathematical foundation. However, it is still the most powerful calculation tool in quantum (and statistical) mechanics. Later, Feynman generalized it to encompass physical fields – which led to his version of quantum electrodynamics (the first prototype of a quantum field theory) – and his Nobel Prize. All his career he consistently distrusted official mathematics and invented his own maths underpinned with a direct physical intuition.

If the story had ended here, we might have said that *visual physical intuition* is leading the way of science. However, the story does not end here. The leading authority in contemporary physics is *Ed Witten*, a physicist who did not get the Nobel Prize, but rather the Fields Medal – together with his 'superstring theory of everything'.⁴⁵ Witten works at the same place where Einstein spent the last 30 years of his life – at the Princeton Institute of Advanced Study. He is dreaming Einstein's dream: a *unified theory of everything*, using the most powerful maths possible. His prophecy, delivered at a turn of the Century, has been: "In the 21st Century, mathematics will be dominated by string theory."

When superstring theory arrived in physics in 1984 as a potential theory of the universe, it was considered by mainstream physicists as little better than religion in terms of constituting a viable, testable theory. In string theory, the fundamental particles were string-like, rather than point particles; the universe had 10 or 11 dimensions, rather than four; and the

Their joined work on gravity is called the *Einstein–Hilbert action*.

⁴⁵Witten joined the 'old Green–Schwarz bosonic string community' after he won his Fields Medal for topological quantum field theory (TQFT)

theory itself existed at an energy so far from earthly energies that it took a leap of enormous faith to imagine the day when an experiment could ever test it. Quite simply, string theory seemed an excessively esoteric pursuit, which it still is.

1.1.8.2 *String Theory in ‘Plain English’*

With modern (super)string theory,⁴⁶ scientists might be on the verge of fulfilling Einstein’s dream: formulating the sought for ‘theory of everything’, which would unite our understanding of the *four fundamental forces* of Nature⁴⁷ into a single equation (like, e.g., Newton, or Einstein, or Schrödinger equation) and explaining the basic nature of matter and energy.

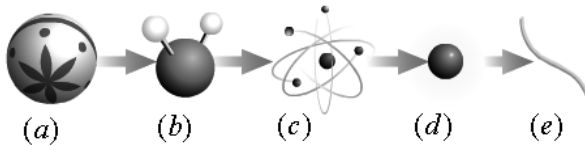


Fig. 1.2 All particles and forces of Nature are supposed to be manifestations of different resonances of tiny 1D strings vibrating in a 10D hyper-space: (a) An ordinary matter; (b) A molecule; (c) An atom (around ten billionths of a centimeter in diameter); (d) A subatomic particle (e.g., proton – around 100,000 times smaller than an atom); (e) A super-string (around 10^{20} times smaller than a proton).

In simplest terms, string theory states that all particles and forces of Nature are manifestations of different resonances of tiny 1-dimensional strings (rather than the zero-dimensional points (particles) that are the basis of the *Standard Model of particle physics*), vibrating in 10 dimensions (see

⁴⁶Recall that ‘superstring’ means ‘supersymmetric string’. The *supersymmetry* (often abbreviated SUSY) is a hypothetical symmetry that relates bosons (particles that transmit forces) and fermions (particles of matter). In supersymmetric theories, every fundamental fermion has a bosonic ‘superpartner’ and vice versa.

⁴⁷Recall that the *four fundamental forces* are: (i) *Gravity* (it describes the attractive force of matter; it is the same force that holds planets and moons in their orbits and keeps our feet on the ground; it is the weakest force of the four by many orders of magnitude); (ii) *Electromagnetism* (it describes how electric and magnetic fields work together; it also makes objects solid; once believed to be two separate forces, could be described by a relatively simple set of *Maxwell equations*); (iii) *Strong nuclear force* (it is responsible for holding the nucleus of atoms together; without it, protons would repel one another so no elements other than hydrogen, which has only one proton, would be able to form); (iv) *Weak nuclear force* (it explains beta decay and the associated radioactivity; it also describes how elementary particles can change into other particles with different energies and masses).

Figure 1.2).

Recall that the *Standard Model* is a theory which describes the strong, weak, and electromagnetic fundamental forces, as well as the fundamental particles that make up all matter. Developed between 1970 and 1973, it is a *quantum field theory*, and consistent with both quantum mechanics and special relativity. The Standard Model contains both *fermionic and bosonic fundamental particles*. Fermions are particles which possess half-integer *spin*, obey the *Fermi–Dirac statistics* and also the *Pauli exclusion principle*, which states that no fermions can share the same quantum state. On the other hand, bosons possess integer spin, obey the *Bose–Einstein statistics*, and do not obey the Pauli exclusion principle. In the Standard Model, the theory of the *electro-weak interaction* (which describes the weak and electromagnetic interactions) is combined with the theory of *quantum chromodynamics*. All of these theories are *gauge theories*,⁴⁸ meaning that they model the forces between fermions by coupling them to bosons which

⁴⁸Recall that the familiar *Maxwell gauge field theory* (or, in the non-Abelian case, *Yang–Mills gauge field theory*) is defined in terms of the fundamental *gauge field* (which geometrically represents a *connection*) $A_\mu = (A_0, \vec{A})$, that is $\mu = 0, 3$. Here A_0 is the scalar potential and \vec{A} is the vector potential. The *Maxwell Lagrangian*

$$\mathcal{L}_M = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - A_\mu J^\mu \quad (1.1)$$

is expressed in terms of the field strength tensor (curvature) $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, and a matter current J^μ that is conserved: $\partial_\mu J^\mu = 0$. This Maxwell Lagrangian is manifestly invariant under the gauge transformation $A_\mu \rightarrow A_\mu + \partial_\mu \Lambda$; and, correspondingly, the classical Euler-Lagrange equations of motion

$$\partial_\mu F^{\mu\nu} = J^\nu \quad (1.2)$$

are gauge invariant. Observe that current conservation $\partial_\nu J^\nu = 0$ follows from the antisymmetry of $F_{\mu\nu}$.

Note that this Maxwell theory could easily be defined in *any* space–time dimension d simply by taking the range of the space–time index μ on the gauge field A_μ to be $\mu = 0, 1, 2, \dots, (d-1)$ in d D space–time. The field strength tensor is still the antisymmetric tensor $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, and the Maxwell Lagrangian (1.1) and the field equations of motion (1.2) do not change their form. The only real difference is that the number of independent fields contained in the field strength tensor $F_{\mu\nu}$ is different in different dimensions. (Since $F_{\mu\nu}$ can be regarded as a $d \times d$ antisymmetric matrix, the number of fields is equal to $\frac{1}{2}d(d-1)$.) So at this level, planar $(2+1)$ D Maxwell theory is quite similar to the familiar $(3+1)$ D Maxwell theory. The main difference is simply that the magnetic field is a (pseudo-) scalar $B = \epsilon^{ij}\partial_i A_j$ in $(2+1)$ D, rather than a (pseudo-) vector $\vec{B} = \vec{\nabla} \times \vec{A}$ in $(3+1)$ D. This is just because in $(2+1)$ D the vector potential \vec{A} is a 2D vector, and the curl in 2D produces a scalar. On the other hand, the electric field $\vec{E} = -\vec{\nabla}A_0 - \dot{\vec{A}}$ is a 2D vector. So the antisymmetric 3×3 field–strength tensor has three nonzero field components: two for the electric field \vec{E} and one for the magnetic field B .

mediate the forces. The Lagrangian of each set of mediating bosons is invariant under a transformation called a *gauge transformation*, so these mediating bosons are referred to as gauge bosons. There are twelve different ‘flavours’ of fermions in the Standard Model. The proton, neutron are

The real novelty of $(2+1)\text{D}$ is that, instead of considering this ‘reduced’ form of Maxwell theory, we can also define a completely different type of gauge theory: a *Chern–Simons gauge theory*. It satisfies the usual *criteria for a sensible gauge theory*: it is Lorentz invariant, gauge invariant, and local. The *Chern–Simons Lagrangian* is (see, e.g., [Dunne (1999)])

$$\mathcal{L}_{\text{CS}} = \frac{\kappa}{2} \epsilon^{\mu\nu\rho} A_\mu \partial_\nu A_\rho - A_\mu J^\mu. \quad (1.3)$$

Two things are important about this Chern–Simons Lagrangian. First, it does not *look* gauge invariant, because it involves the gauge field A_μ itself, rather than just the (manifestly gauge invariant) field strength $F_{\mu\nu}$. Nevertheless, under a gauge transformation, the Chern–Simons Lagrangian changes by a total space–time derivative

$$\delta \mathcal{L}_{\text{CS}} = \frac{\kappa}{2} \partial_\mu (\lambda \epsilon^{\mu\nu\rho} \partial_\nu A_\rho). \quad (1.4)$$

Therefore, if we can neglect boundary terms then the corresponding *Chern–Simons action*,

$$S_{\text{CS}} = \int d^3x \mathcal{L}_{\text{CS}},$$

is gauge invariant. This is reflected in the fact that the classical Euler–Lagrange equations

$$\frac{\kappa}{2} \epsilon^{\mu\nu\rho} F_{\nu\rho} = J^\mu, \quad \text{or equivalently} \quad F_{\mu\nu} = \frac{1}{\kappa} \epsilon_{\mu\nu\rho} J^\rho, \quad (1.5)$$

are clearly gauge invariant. Note that the *Bianchi identity*, $\epsilon^{\mu\nu\rho} \partial_\mu F_{\nu\rho} = 0$, is compatible with the *current conservation*: $\partial_\mu J^\mu = 0$, which follows from the Noether Theorem. A second important feature of the Chern–Simons Lagrangian (1.3) is that it is *first-order* in space–time derivatives. This makes the canonical structure of these theories significantly different from that of Maxwell theory. A related property is that the Chern–Simons Lagrangian is particular to $(2+1)\text{D}$, in the sense that we cannot write down such a term in $(3+1)\text{D}$ – the indices simply do not match up. Actually, it is possible to write down a ‘Chern–Simons theory’ in any odd space–time dimension (for example, the Chern–Simons Lagrangian in 5D space–time is $\mathcal{L} = \epsilon^{\mu\nu\rho\sigma\tau} A_\mu \partial_\nu A_\rho \partial_\sigma A_\tau$), but it is only in $(2+1)\text{D}$ that the Lagrangian is quadratic in the gauge field.

Recently, increasingly popular has become *Seiberg–Witten gauge theory*. It refers to a set of calculations that determine the low–energy physics, namely the *moduli space* and the masses of electrically and magnetically charged supersymmetric particles as a function of the moduli space. This is possible and nontrivial in gauge theory with $N = 2$ *extended supersymmetry*, by combining the fact that various parameters of the Lagrangian are holomorphic functions (a consequence of supersymmetry) and the known behavior of the theory in the classical limit. The extended supersymmetry is *supersymmetry* whose infinitesimal generators Q_i^α carry not only a spinor index α , but also an additional index $i = 1, 2, \dots$ The more extended supersymmetry is, the more it constrains physical observables and parameters. Only the minimal (un–extended) supersymmetry is a realistic conjecture for particle physics, but extended supersymmetry is very important for analysis of mathematical properties of *quantum field theory* and *superstring theory*.

made up of two of these: the up quark and down quark, bound together by the strong nuclear force. Together with the electron (bound to the nucleus in atoms by the electromagnetic force), those fermions constitute the vast majority of everyday matter. To date, almost all experimental tests of the three forces described by the Standard Model have agreed with its predictions. However, the Standard Model is not a complete theory of fundamental interactions, primarily because it does not describe the gravitational force.

For this reason, string theories are able to avoid problems associated with the presence of point-like particles in a physical theory. The basic idea is that the fundamental constituents of Nature are *strings of energy* of the *Planck length* (around 10^{-35} m), which *vibrate at specific resonant frequencies* (modes). Another key claim of the theory is that no measurable differences can be detected between strings that wrap around dimensions smaller than themselves and those that move along larger dimensions (i.e., physical processes in a dimension of size R match those in a dimension of size $1/R$). Singularities are avoided because the observed consequences of ‘big crunches’ never reach zero size. In fact, should the universe begin a ‘big crunch’ sort of process, string theory dictates that the universe could never be smaller than the size of a string, at which point it would actually begin expanding.

Recently, physicists have been exploring the possibility that the strings are actually membranes, that is strings with 2 or more dimensions (membranes are referred to as p -branes, where p is the number of dimensions, see Figure 1.3). Every p -brane sweeps out a $(p + 1)$ -dimensional *world-volume* as it propagates through space-time. A special class of p -branes are the so-called *D-branes*, named for the mathematician Johann Dirichlet.⁴⁹ D-branes are typically classified by their dimension, which is indicated by a number written after the D: a D0-brane is a single point, a D1-brane is a line (sometimes called a ‘D-string’), a D2-brane is a plane, and a D25-brane fills the highest-dimensional space considered in old bosonic string

⁴⁹Recall that *Dirichlet boundary conditions* have long been used in the study of fluids and potential theory, where they involve specifying some quantity all along a boundary. In *fluid dynamics*, fixing a Dirichlet boundary condition could mean assigning a known fluid velocity to all points on a surface; when studying *electrostatics*, one may establish Dirichlet boundary conditions by fixing the voltage to known values at particular locations, like the surfaces of conductors. In either case, the locations at which values are specified is called a *D-brane*. These constructions take on special importance in string theory, because open strings must have their endpoints attached to D-branes.

theory.⁵⁰

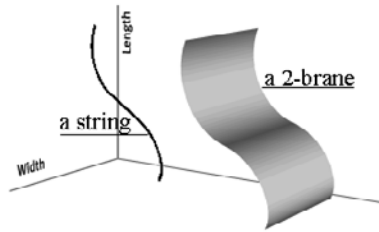


Fig. 1.3 Visualizing strings and p -branes.

According to superstring theory, all the different types of elementary particles can be derived from only five types of interactions between just two different states of strings, *open* and *closed*: (i) an open string can split to create two smaller open strings (see Figure 1.4); (ii) a closed string can split to create two smaller closed strings; (iii) an open string can form both a new open and a new closed string; (iv) two open strings can collide and create two new open strings; (v) an open string can join its ends to become a closed string. All the forces and particles of Nature are just *different modes of vibrating strings* (somewhat like vibrating strings on string instruments to produce a music: different strings have different frequencies that sound as different notes and combining several strings gives chords). For example, gravity is caused by the lowest vibratory mode of a circular string. Higher frequencies and different interactions of superstrings create different forms of matter and energy.

⁵⁰The central idea of the so-called *brane-world scenario* is that our visible 3D universe is entirely restricted to a D3-brane embedded in a higher-dimensional space-time, called the *bulk*. The additional dimensions may be taken to be compact, in which case the observed universe contains the extra dimensions, and then no reference to the bulk is appropriate in this context. In the bulk model, other branes may be moving through this bulk. Interactions with the bulk, and possibly with other branes, can influence our brane and thus introduce effects not seen in more standard cosmological models. As one of its attractive features, the model can ‘explain’ the weakness of gravity relative to the other fundamental forces of nature. In the brane picture, the other three forces (electromagnetism and the weak and strong nuclear forces) are localized on the brane, but gravity has no such constraint and so much of its attractive power ‘leaks’ into the bulk. As a consequence, the force of gravity should appear significantly stronger on small (sub-millimetre) scales, where less gravitational force has ‘leaked’. Various experiments are currently underway to test this. For example, in a particle accelerator, if a graviton were to be discovered and then observed to suddenly disappear, it might be assumed that the graviton ‘leaked’ into the bulk.

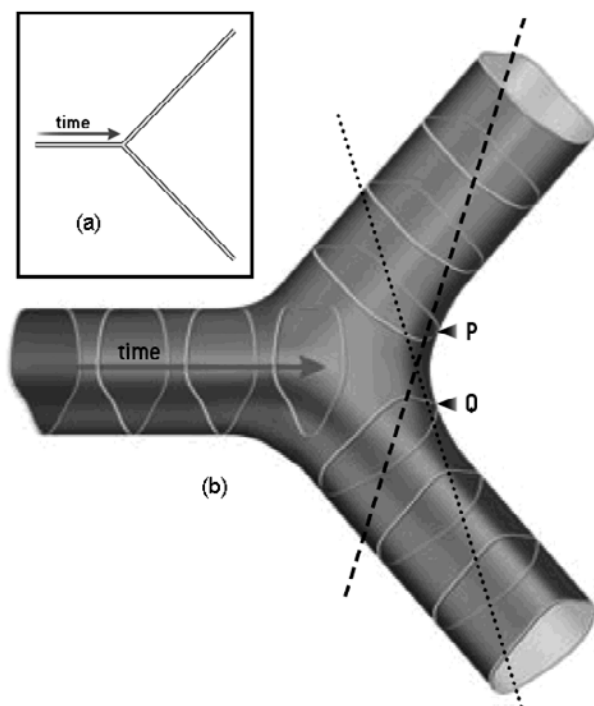


Fig. 1.4 An elementary particle split (a) and string split (b). When a single elementary particle splits in two particles, it occurs at a definite moment in space-time. On the other hand, when a string splits into two strings, different observers will disagree about when and where this occurred. A relativistic observer who considers the dotted line to be a surface of constant time believes the string broke at the space-time point P while another observer who considers the dashed line to be a surface of constant time believes the string broke at Q.

String theory is a possible solution of the *core quantum gravity problem*, and in addition to gravity it can naturally describe interactions similar to electromagnetism and the other forces of nature. Superstring theories include *fermions*, the building blocks of matter, and incorporate the so-called *supersymmetry*.⁵¹ It is not yet known whether string theory will be able

⁵¹In a world based on *supersymmetry*, when a particle moves in space, it also can vibrate in the new fermionic dimensions. This new kind of vibration produces a ‘cousin’ or ‘superpartner’ for every elementary particle that has the same electric charge but differs in other properties such as *spin*. Supersymmetric theories make detailed predictions about how superpartners will behave. To confirm supersymmetry, scientists would like to produce and study the new supersymmetric particles. The crucial step is building a particle accelerator that achieves high enough energies. At present, the

to describe a universe with the precise collection of forces and matter that is observed, nor how much freedom to choose those details that the theory will allow. String theory as a whole has not yet made falsifiable predictions that would allow it to be experimentally tested, though various special corners of the theory are accessible to planned observations and experiments. Work on string theory has led to advances in both mathematics (mainly in differential and algebraic geometry) and physics (supersymmetric gauge theories).⁵²

Historically, string theory was originally invented to explain peculiarities

highest-energy particle accelerator is the Tevatron at Fermilab near Chicago. There, protons and antiprotons collide with an energy nearly 2,000 times the rest energy of an individual proton (given by Einsteins well-known formula $E = mc^2$). Earlier in this decade, physicists capitalized on Tevatron's unsurpassed energy in their discovery of the top quark, the heaviest known elementary particle. After a shutdown of several years, the Tevatron resumed operation in 2001 with even more intense particle beams. In 2007, the available energies will make a 'quantum jump' when the European Laboratory for Particle Physics, or CERN (located near Geneva, Switzerland) turns on the Large Hadron Collider (LHC). The LHC should reach energies 15,000 times the proton rest energy. The LHC is a multi-billion dollar international project, funded mainly by European countries with substantial contributions from the United States, Japan, and other countries.

⁵²Recall that *gauge theories* are a class of physical theories based on the idea that symmetry transformations can be performed locally as well as globally. *Yang-Mills theory* is a particular example of gauge theories with non-Abelian symmetry groups specified by the *Yang-Mills action*. For example, the Yang-Mills action for the $O(n)$ gauge theory for a set of n non-interacting scalar fields φ_i , with equal masses m is

$$S = \int \left(\sum_{i=1}^n \frac{1}{2} \partial_\mu \varphi_i \partial^\mu \varphi_i - \frac{1}{2} m^2 \varphi_i^2 \right) d^4 x.$$

Other gauge theories with a non-Abelian gauge symmetry also exist, e.g., the Chern-Simons model. Most physical theories are described by Lagrangians which are invariant under certain transformations, when the transformations are identically performed at every space-time point-they have global symmetries. Gauge theory extends this idea by requiring that the Lagrangians must possess local symmetries as well-it should be possible to perform these symmetry transformations in a particular region of space-time without affecting what happens in another region. This requirement is a generalized version of the equivalence principle of general relativity. Gauge symmetries reflect a redundancy in the description of a system. The importance of gauge theories for physics stems from the tremendous success of the mathematical formalism in providing a unified framework to describe the quantum field theories of electromagnetism, the weak force and the strong force. This theory, known as the Standard Model (see footnote 5), accurately describes experimental predictions regarding three of the four fundamental forces of nature, and is a gauge theory with the gauge group $SU(3) \times SU(2) \times U(1)$. Modern theories like string theory, as well as some formulations of general relativity, are, in one way or another, gauge theories. Sometimes, the term gauge symmetry is used in a more general sense to include any local symmetry, like for example, *diffeomorphisms*.

of hadron (subatomic particle which experiences the strong nuclear force) behavior. In particle-accelerator experiments, physicists observed that the spin of a hadron is never larger than a certain multiple of the square of its energy. No simple model of the hadron, such as picturing it as a set of smaller particles held together by spring-like forces, was able to explain these relationships. In 1968, theoretical physicist Gabriele Veneziano was trying to understand the strong nuclear force when he made a startling discovery. He found that a 200-year-old *Euler beta function* perfectly matched modern data on the strong force. Veneziano applied the Euler beta function to the strong force, but no one could explain why it worked.

In 1970, Yoichiro Nambu, Holger Bech Nielsen, and Leonard Susskind presented a physical explanation for Euler's strictly theoretical formula. By representing nuclear forces as vibrating, 1D strings, these physicists showed how Euler's function accurately described those forces. But even after physicists understood the physical explanation for Veneziano's insight, the string description of the strong force made many predictions that directly contradicted experimental findings. The scientific community soon lost interest in string theory, and the Standard Model, with its particles and fields, remained un-threatened.

Then, in 1974, John Schwarz and Joel Scherk studied the messenger-like patterns of string vibration and found that their properties exactly matched those of the gravitational force's hypothetical messenger particle - the *graviton*. They argued that string theory had failed to catch on because physicists had underestimated its scope. This led to the development of *bosonic string theory*, which is still the version first taught to many students. The original need for a viable theory of hadrons has been fulfilled by *quantum chromodynamics* (QCD), the theory of Gell-Mann's *quarks* and their interactions. It is now hoped that string theory (or some descendant of it) will provide a fundamental understanding of the quarks themselves.

Bosonic string theory is formulated in terms of the so-called *Polyakov action*, a mathematical quantity which can be used to predict how strings move through space and time. By applying the ideas of quantum mechanics to the Polyakov action - a procedure known as *quantization* - one can deduce that each string can vibrate in many different ways, and that each vibrational state appears to be a different particle. The mass the particle has, and the fashion with which it can interact, are determined by the way the string vibrates - in essence, by the 'note' which the string sounds. The scale of notes, each corresponding to a different kind of particle, is termed the *spectrum* of the theory. These early models included both open strings,

which have two distinct endpoints, and closed strings, where the endpoints are joined to make a complete loop. The two types of string behave in slightly different ways, yielding two spectra. Not all modern string theories use both types; some incorporate only the closed variety. However, the bosonic theory has problems. Most importantly, the theory has a fundamental instability, believed to result in the decay of space-time itself. Additionally, as the name implies, the spectrum of particles contains only bosons, particles like the photon which obey particular rules of behavior. While bosons are a critical ingredient of the Universe, they are not its only constituents. Investigating how a string theory may include fermions in its spectrum led to supersymmetry, a mathematical relation between bosons and fermions which is now an independent area of study. String theories which include fermionic vibrations are now known as superstring theories; several different kinds have been described.

Roughly between 1984 and 1986, physicists realized that string theory could describe all elementary particles and interactions between them, and hundreds of them started to work on string theory as the most promising idea to unify theories of physics. This so-called *first superstring revolution* was started by a discovery of *anomaly* cancellation in type I string theory by Michael Green and John Schwarz in 1984. The anomaly is cancelled due to the *Green-Schwarz mechanism*. Several other ground-breaking discoveries, such as the heterotic string, were made in 1985.

Note that in the type IIA and type IIB string theories closed strings are allowed to move everywhere throughout the 10D space-time (called the *bulk*), while open strings have their ends attached to *D-branes*, which are membranes of lower dimensionality (their dimension is odd - 1,3,5,7 or 9 - in type IIA and even - 0,2,4,6 or 8 - in type IIB, including the time direction).

While understanding the details of string and superstring theories requires considerable geometrical sophistication, some qualitative properties of quantum strings can be understood in a fairly intuitive fashion. For example, quantum strings have tension, much like regular strings made of twine; this tension is considered a fundamental parameter of the theory. The tension of a quantum string is closely related to its size. Consider a closed loop of string, left to move through space without external forces. Its tension will tend to contract it into a smaller and smaller loop. Classical intuition suggests that it might shrink to a single point, but this would violate *Heisenberg's uncertainty principle*. The characteristic size of the string loop will be a balance between the tension force, acting to make it

Contemporary String Theories		
Type	Dim	Details
Bosonic	26	Only <i>bosons</i> , no <i>fermions</i> means only forces, no matter, with both open and closed strings; major flaw: a <i>particle</i> with imaginary <i>mass</i> , called the <i>tachyon</i> , representing an instability in the theory
I	10	<i>Supersymmetry</i> between forces and matter, with both open and closed strings, no <i>tachyon</i> , group symmetry is $SO(32)$
IIA	10	<i>Supersymmetry</i> between forces and matter, with closed strings and open strings bound to <i>D-branes</i> , no <i>tachyon</i> , massless <i>fermions</i> spin both ways (nonchiral)
IIB	10	<i>Supersymmetry</i> between forces and matter, with closed strings and open strings bound to <i>D-branes</i> , no <i>tachyon</i> , massless <i>fermions</i> only spin one way (chiral)
HO	10	<i>Supersymmetry</i> between forces and matter, with closed strings only, no <i>tachyon</i> , <i>heterotic</i> , meaning right moving and left moving strings differ, group symmetry is $SO(32)$
HE	10	<i>Supersymmetry</i> between forces and matter, with closed strings only, no <i>tachyon</i> , <i>heterotic</i> , meaning right moving and left moving strings differ, group symmetry is $E_8 \times E_8$

small, and the uncertainty effect, which keeps it ‘stretched’. Consequently, the minimum size of a string must be related to the string tension.

Before the 1990s, string theorists believed that there were five distinct superstring theories: *type I*, *types IIA* and *IIB*, and the two *heterotic string* theories ($SO(32)$ and $E_8 \times E_8$). The thinking was that out of these five candidate theories, only one was the actual correct *theory of everything*, and that theory was the theory whose low energy limit, with ten dimensions spacetime *compactified* down to four, matched the physics observed in our world today. But now it is known that this naïve picture was wrong, and that the five superstring theories are connected to one another as if they are each a special case of some more fundamental theory, of which there is only one. These theories are related by transformations that are called dualities. If two theories are related by a duality transformation, it means that the

first theory can be transformed in some way so that it ends up looking just like the second theory. The two theories are then said to be dual to one another under that kind of transformation. Put differently, the two theories are two different mathematical descriptions of the same phenomena. These dualities link quantities that were also thought to be separate. Large and small distance scales, strong and weak coupling strengths – these quantities have always marked very distinct limits of behavior of a physical system, in both classical *field theory* and quantum *particle physics*. But strings can obscure the difference between large and small, strong and weak, and this is how these five very different theories end up being related.

This type of duality is called *T-duality*. T-duality relates type IIA superstring theory to type IIB superstring theory. That means if we take type IIA and Type IIB theory and ‘compactify’ them both on a circle, then switching the momentum and winding modes, and switching the distance scale, changes one theory into the other. The same is also true for the two heterotic theories. T-duality also relates type I superstring theory to both type IIA and type IIB superstring theories with certain boundary conditions (termed ‘orientifold’). Formally, the location of the string on the circle is described by two fields living on it, one which is left-moving and another which is right-moving. The movement of the string center (and hence its momentum) is related to the sum of the fields, while the string stretch (and hence its winding number) is related to their difference. T-duality can be formally described by taking the left-moving field to minus itself, so that the sum and the difference are interchanged, leading to switching of momentum and winding.

On the other hand, every force has a *coupling constant*, which is a measure of its strength, and determines the chances of one particle to emit or receive another particle. For electromagnetism, the coupling constant is proportional to the square of the electric charge. When physicists study the quantum behavior of electromagnetism, they can’t solve the whole theory exactly, because every particle may emit and receive many other particles, which may also do the same, endlessly. So events of emission and reception are considered as perturbations and are dealt with by a series of approximations, first assuming there is only one such event, then correcting the result for allowing two such events, etc (this method is called Perturbation theory. This is a reasonable approximation only if the coupling constant is small, which is the case for electromagnetism. But if the coupling constant gets large, that method of calculation breaks down, and the little pieces become worthless as an approximation to the real physics. This can also

happen in string theory. String theories have a *string coupling constant*. But unlike in particle theories, the string coupling constant is not just a number, but depends on one of the oscillation modes of the string, called the *dilaton*. Exchanging the dilaton field with minus itself exchanges a very large coupling constant with a very small one. This symmetry is called *S-duality*. If two string theories are related by S-duality, then one theory with a strong coupling constant is the same as the other theory with weak coupling constant. The theory with strong coupling cannot be understood by means of perturbation theory, but the theory with weak coupling can. So if the two theories are related by S-duality, then we just need to understand the weak theory, and that is equivalent to understanding the strong theory. Superstring theories related by S-duality are: type I superstring theory with heterotic $SO(32)$ superstring theory, and type IIB theory with itself.

Around 1995, Edward Witten and others found strong evidence that the different superstring theories were different limits of a new 11D theory called M-theory. With the discovery of M-theory, an extra dimension appeared and the fundamental string of string theory became a 2-dimensional membrane called an M2-brane (or supermembrane). Its magnetic dual is an M5-brane. The various branes of string theory are thought to be related to these higher dimensional M-branes wrapped on various cycles. These discoveries sparked the so-called *second superstring revolution*.

One intriguing feature of string theory is that it *predicts* the number of dimensions which the universe should possess. Nothing in Maxwell's theory of electromagnetism, or Einstein's theory of relativity, makes this kind of prediction; these theories require physicists to insert the number of dimensions 'by hand'. The first person to add a fifth dimension to Einstein's four space-time dimensions was German mathematician Theodor Kaluza in 1919. The reason for the un-observability of the fifth dimension (its compactness) was suggested by Swedish physicist Oskar Klein in 1926. Today, this is called the 5D *Kaluza-Klein theory*.

Instead, string theory allows one to compute the number of spacetime dimensions from first principles. Technically, this happens because for a different number of dimensions, the theory has a gauge anomaly. This can be understood by noting that in a consistent theory which includes a photon (technically, a particle carrying a force related to an unbroken gauge symmetry), it must be massless. The mass of the photon which is predicted by string theory depends on the energy of the string mode which represents the photon. This energy includes a contribution from

the *Casimir effect*, namely from *quantum fluctuations in the string*. The size of this contribution depends on the number of dimensions since for a larger number of dimensions, there are more possible fluctuations in the string position. Therefore, the photon will be massless – and the theory consistent – only for a particular number of dimensions.

The only problem is that when the calculation is done, the universe's dimensionality is not four as one may expect (three axes of space and one of time), but 26. More precisely, bosonic string theories are 26D, while superstring and M-theories turn out to involve 10 and 11 dimensions, respectively. In bosonic string theories, the 26 dimensions come from the *Polyakov equation*. However, these results appear to contradict the observed four dimensional space–time.

Two different ways have been proposed to solve this apparent contradiction. The first is to *compactify the extra dimensions*; i.e., the 6 or 7 extra dimensions are so small as to be undetectable in our phenomenal experience. The 6D model's resolution is achieved with the so-called *Calabi–Yau manifolds* (see Figure 1.5). In 7D, they are termed G_2 –manifolds. Essentially these extra dimensions are compactified by causing them to loop back upon themselves. A standard analogy for this is to consider multidimensional space as a garden hose. If the hose is viewed from a sufficient distance, it appears to have only one dimension, its length. Indeed, think of a ball small enough to enter the hose but not too small. Throwing such a ball inside the hose, the ball would move more or less in one dimension; in any experiment we make by throwing such balls in the hose, the only important movement will be one-dimensional, that is, along the hose. However, as one approaches the hose, one discovers that it contains a second dimension, its circumference. Thus, a ant crawling inside it would move in two dimensions (and a fly flying in it would move in three dimensions). This 'extra dimension' is only visible within a relatively close range to the hose, or if one 'throws in' small enough objects. Similarly, the extra compact dimensions are only visible at extremely small distances, or by experimenting with particles with extremely small wave lengths (of the order of the compact dimension's radius), which in quantum mechanics means very high energies. Another possibility is that we are stuck in a 3+1 dimensional (i.e., three spatial dimensions plus one time dimension) subspace of the full universe. This subspace is supposed to be a D-brane, hence this is known as a *brane-world theory*. In either case, gravity acting in the hidden dimensions affects other non-gravitational forces such as electromagnetism. In principle, therefore, it is possible to deduce the nature of those extra dimensions

by requiring consistency with the Standard Model, but this is not yet a practical possibility. It is also possible to extract information regarding the hidden dimensions by precision tests of gravity, but so far these have only put upper limitations on the size of such hidden dimensions.

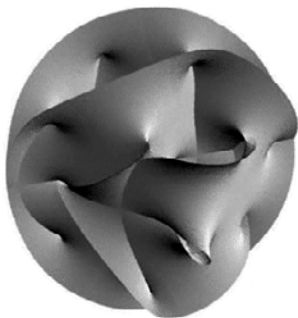


Fig. 1.5 Calabi-Yau manifold – a 3D projection created using MathematicaTM.

For popular expose on string theory, see [Witten (2002); Greene (2000)], while the main textbook is still [Green *et. al.* (1987)].

1.2 Application: Paradigm of Differential-Geometric Modelling of Dynamical Systems

In this section we give a paradigm of differential-geometric modelling and analysis of complex dynamical systems (see [Ivancevic and Ivancevic (2006)] for more background details). This is essentially a *recipe* how to develop a *covariant formalism on smooth manifolds*, given a certain physical, or bio-physical, or psycho-physical, or socio-physical system, here labelled by a generic name: ‘physical situation’. We present this recipe in the form of the following five-step algorithm.

(I) So let’s start: given a certain physical situation, the first step in its predictive modelling and analysis, that is, in applying a powerful differential-geometric machinery to it, is to associate with this situation *two* independent coordinate systems, constituting two independent smooth Riemannian manifolds. Let us denote these two coordinate systems and their respective manifolds as:

- *Internal coordinates:* $x^i = x^i(t)$, ($i = 1, \dots, m$), constituting the mD *internal configuration manifold:* $M^m \equiv \{x^i\}$; and

- *External coordinates:* $y^e = y^e(t)$, ($e = 1, \dots, n$), constituting the nD *external configuration manifold:* $N^n \equiv \{y^e\}$.

The main example that we have in mind is a standard robotic or bio-dynamic (loco)motion system, in which x^i denote internal joint coordinates, while y^e denote external Cartesian coordinates of segmental centers of mass. However, we believe that such developed methodology can fit a generic physical situation.

Therefore, in this first, engineering step (I) of our differential-geometric modelling, we associate to the given natural system, not one but two different and independent smooth configuration manifolds, somewhat like viewing from two different satellites a certain place on Earth with a football game playing in it.

(II) Once that we have precisely defined two smooth manifolds, as two independent views on the given physical situation, we can apply our differential-geometric modelling to it and give it a natural physical interpretation. More precisely, once we have two smooth Riemannian manifolds, $M^m \equiv \{x^i\}$ and $N^n \equiv \{y^e\}$, we can formulate two smooth maps between them:⁵³

$$f : N \rightarrow M, \text{ given by coordinate transformation: } x^i = f^i(y^e), \quad (1.6)$$

and

$$g : M \rightarrow N, \text{ given by coordinate transformation: } y^e = g^e(x^i). \quad (1.7)$$

If the Jacobian matrices of these two maps are nonsingular (regular), that is if their Jacobian determinants are nonzero, then these two maps are mutually inverse, $f = g^{-1}$, and they represent standard *forward and inverse kinematics*.

(III) Although, maps f and g define some completely general nonlinear coordinate (functional) transformations, which are even unknown at the moment, there is something linear and simple that we know about them (from calculus). Namely, the corresponding infinitesimal transformations are linear and homogenous: from (1.6) we have (applying everywhere Einstein's summation convention over repeated indices)

$$dx^i = \frac{\partial f^i}{\partial y^e} dy^e, \quad (1.8)$$

⁵³This obviously means that we are working in the category of smooth manifolds.

while from (1.7) we have

$$dy^e = \frac{\partial g^e}{\partial x^i} dx^i. \quad (1.9)$$

Furthermore, (1.8) implies the linear and homogenous transformation of *internal velocities*,

$$v^i \equiv \dot{x}^i = \frac{\partial f^i}{\partial y^e} \dot{y}^e, \quad (1.10)$$

while (1.9) implies the linear and homogenous transformation of *external velocities*,

$$u^e \equiv \dot{y}^e = \frac{\partial g^e}{\partial x^i} \dot{x}^i. \quad (1.11)$$

In this way, we have defined *two velocity vector-fields*, the internal one: $v^i = v^i(x^i, t)$ and the external one: $u^e = u^e(y^e, t)$, given respectively by the two nonlinear systems of ODEs, (1.10) and (1.11).⁵⁴

(IV) The next step in our differential-geometrical modelling/analysis is to define second derivatives of the manifold maps f and g , that is the two *acceleration vector-fields*, which we will denote by $a^i = a^i(x^i, \dot{x}^i, t)$ and $w^e = w^e(y^e, \dot{y}^e, t)$, respectively. However, unlike simple physics in linear Euclidean spaces, these two acceleration vector-fields on manifolds M and N are not the simple time derivatives of the corresponding velocity vector-fields ($a^i \neq \dot{v}^i$ and $w^e \neq \dot{u}^e$), due to the existence of the *Levi-Civita connections* ∇_M and ∇_N on both M and N . Properly defined, these two acceleration vector-fields respectively read:

$$a^i = \dot{v}^i + \Gamma_{jk}^i v^j v^k = \ddot{x}^i + \Gamma_{jk}^i \dot{x}^j \dot{x}^k, \quad \text{and} \quad (1.12)$$

$$w^e = \dot{u}^e + \Gamma_{hl}^e u^h u^l = \ddot{y}^e + \Gamma_{hl}^e \dot{y}^h \dot{y}^l, \quad (1.13)$$

where Γ_{jk}^i and Γ_{hl}^e denote the (second-order) *Christoffel symbols* of the connections ∇_M and ∇_N .

Therefore, in the step (III) we gave the first-level model of our physical situation in the form of two ordinary vector-fields, the first-order vector-fields (1.10) and (1.11). For some simple situations (e.g., modelling ecological systems), we could stop at this modelling level. Using physical terminology we call them velocity vector-fields. Following this, in the step (IV) we have defined the two second-order vector-fields (1.12) and (1.13), as

⁵⁴Although transformations of differentials and associated velocities are linear and homogeneous, the systems of ODE's define nonlinear vector-fields, as they include Jacobian (functional) matrices.

a connection–base derivations of the previously defined first–order vector–fields. Using physical terminology, we call them ‘acceleration vector–fields’.

(V) Finally, following our generic physical terminology, as a natural next step we would expect to define some kind of generic Newton–Maxwell force–fields. And we can actually do this, with a little surprise that individual forces involved in the two force–fields will not be vectors, but rather the dual objects called 1–forms (or, 1D differential forms). Formally, we define the two *covariant force–fields* as

$$F_i = \mathbf{m} g_{ij} a^j = \mathbf{m} g_{ij} (\dot{v}^j + \Gamma_{ik}^j v^i v^k) = \mathbf{m} g_{ij} (\ddot{x}^j + \Gamma_{ik}^j \dot{x}^i \dot{x}^k), \text{ and } (1.14)$$

$$G_e = \mathbf{m} g_{eh} w^h = \mathbf{m} g_{eh} (\dot{u}^h + \Gamma_{el}^h u^e u^l) = \mathbf{m} g_{eh} (\dot{y}^h + \Gamma_{el}^h \dot{y}^e \dot{y}^l), \quad (1.15)$$

where \mathbf{m} is the mass of each single segment (unique, for simplicity), while $g_{ij} = g_{ij}^M$ and $g_{eh} = g_{eh}^N$ are the two *Riemannian metric tensors* corresponding to the manifolds M and N . The two force–fields, F_i defined by (1.14) and G_e defined by (1.15), are generic force–fields corresponding to the manifolds M and N , which represent the *material cause* for the given physical situation. Recall that they can be physical, bio–physical, psycho–physical or socio–physical force–fields. Physically speaking, they are the *generators* of the corresponding dynamics and kinematics.

Main geometrical relations behind this fundamental paradigm, forming the so–called *covariant force functor*, are depicted in Figure 1.6.

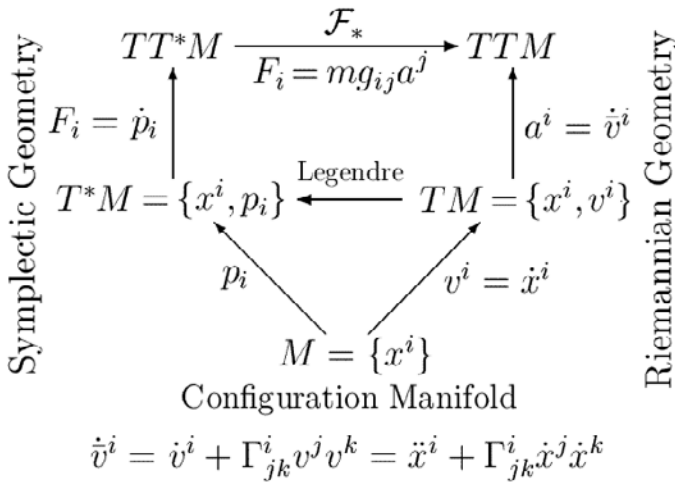


Fig. 1.6 The covariant force functor, including the main relations used by differential–geometric modelling.

This page intentionally left blank

Chapter 2

Technical Preliminaries: Tensors, Actions and Functors

2.1 Tensors: Local Machinery of Differential Geometry

Physical and engineering laws must be independent of any particular coordinate systems used in describing them mathematically, if they are to be valid. In other words, all physical and engineering equations need to be tensorial or *covariant*. Therefore, for the reference purpose, in this section, we give the basic formulas from the standard tensor calculus, which is used throughout the text. The basic notational convention used in tensor calculus is *Einstein's summation convention* over repeated indices. More on this subject can be found in any standard textbook on mathematical methods for scientists and engineers, or mathematical physics (we recommend [Misner *et al.* (1973)]).

2.1.1 Transformation of Coordinates and Elementary Tensors

To introduce tensors, consider a standard linear n D matrix system, $\mathbf{Ax} = \mathbf{b}$. It can be rewritten in the so-called *covariant form* as

$$a_{ij}x^j = b_i, \quad (i, j = 1, \dots, n). \quad (2.1)$$

Here, i is a *free index* and j is a *dummy index* to be summed upon, so the expansion of (2.1) gives

$$\begin{aligned} a_{11}x^1 + a_{12}x^2 + \dots + a_{1n}x^n &= b_1, \\ a_{21}x^1 + a_{22}x^2 + \dots + a_{2n}x^n &= b_2, \\ &\dots \\ a_{n1}x^1 + a_{n2}x^2 + \dots + a_{nn}x^n &= b_n, \end{aligned}$$

as expected from the original matrix form $\mathbf{Ax} = \mathbf{b}$. This indicial notation can be more useful than the matrix one, like e.g., in computer science, where indices would represent loop variables. However, the full potential of tensor analysis is to deal with nonlinear multivariate systems, which are untractable by linear matrix algebra and analysis. The core of this *nonlinear multivariate analysis* is *general functional transformation*.

2.1.1.1 Transformation of Coordinates

Suppose that we have two sets of curvilinear coordinates that are single-valued, continuous and smooth functions of time, $x^j = x^j(t)$, ($j = 1, \dots, m$) and $\bar{x}^i = \bar{x}^i(t)$, ($i = 1, \dots, n$), respectively, representing trajectories of motion of some physical or engineering system. Then a general $(m \times n)$ D transformation (i.e., a nonlinear map) $x^j \mapsto \bar{x}^i$ is defined by the set of transformation equations

$$\bar{x}^i = \bar{x}^i(x^j), \quad (i = 1, \dots, n; j = 1, \dots, m). \quad (2.2)$$

In case of the square transformation, $m = n$, we can freely exchange the indices, like e.g., in general relativity theory. On the other hand, in the general case of rectangular transformation, $m \neq n$, like e.g., in robotics, and we need to take care of these ‘free’ indices.

Now, if the *Jacobian determinant* of this *coordinate transformation* is different from zero,

$$\left| \frac{\partial \bar{x}^i}{\partial x^j} \right| \neq 0,$$

then the transformation (2.2) is reversible and the inverse transformation,

$$x^j = x^j(\bar{x}^i),$$

exists as well. Finding the inverse transformation is the problem of matrix inverse: in case of the square matrix it is well defined, although the inverse might not exist if the matrix is singular. However, in case of the square matrix, its proper inverse does not exist, and the only tool that we are left with is the so-called *Moore–Penrose pseudoinverse*, which gives an optimal solution (in the least-squares sense) of an overdetermined system of equations. Every (overdetermined) rectangular coordinate transformation induces a *redundant system*.

For example, in Euclidean 3D space \mathbb{R}^3 , transformation from Cartesian coordinates $y^k = \{x, y, z\}$ into spherical coordinates $x^i = \{\rho, \theta, \varphi\}$ is given

by

$$y^1 = x^1 \cos x^2 \cos x^3, \quad y^2 = x^1 \sin x^2 \cos x^3, \quad y^3 = x^1 \sin x^3, \quad (2.3)$$

with the Jacobian matrix given by

$$\left(\frac{\partial y^k}{\partial x^i} \right) = \begin{pmatrix} \cos x^2 \cos x^3 & -x^1 \sin x^2 \cos x^3 & -x^1 \cos x^2 \sin x^3 \\ \sin x^2 \cos x^3 & x^1 \cos x^2 \cos x^3 & -x^1 \sin x^2 \sin x^3 \\ \sin x^3 & 0 & x^1 \cos x^3 \end{pmatrix} \quad (2.4)$$

and the corresponding Jacobian determinant, $\left| \frac{\partial y^k}{\partial x^i} \right| = (x^1)^2 \cos x^3$.

An inverse transform is given by

$$x^1 = \sqrt{(y^1)^2 + (y^2)^2 + (y^3)^2}, \quad x^2 = \arctan \left(\frac{y^2}{y^1} \right),$$

$$x^3 = \arctan \left(\frac{y^3}{\sqrt{(y^1)^2 + (y^2)^2}} \right), \quad \text{with} \quad \left| \frac{\partial x^i}{\partial y^k} \right| = \frac{1}{(x^1)^2 \cos x^3}.$$

As an important engineering (robotic) example, we have a rectangular transformation from 6 DOF external, end-effector (e.g., hand) coordinates, into n DOF internal, joint-angle coordinates. In most cases this is a redundant manipulator system, with infinite number of possible joint trajectories.

2.1.1.2 Scalar Invariants

A *scalar invariant* (or, a zeroth order tensor) with respect to the transformation (2.2) is the quantity $\varphi = \varphi(t)$ defined as

$$\varphi(x^i) = \bar{\varphi}(\bar{x}^i),$$

which does not change at all under the coordinate transformation. In other words, φ is *invariant* under (2.2). For example, biodynamic examples of scalar invariants include various energies (kinetic, potential, biochemical, mental) with the corresponding kinds of work, as well as related thermodynamic quantities (free energy, temperature, entropy, etc.).

2.1.1.3 Vectors and Covectors

Any geometrical object $v^i = v^i(t)$ that under the coordinate transformation (2.2) transforms as

$$\bar{v}^i = v^j \frac{\partial \bar{x}^i}{\partial x^j}, \quad (\text{remember, summing upon } j\text{-index}),$$

represents a *vector*, traditionally called a *contravariant vector*, or, a first-order contravariant tensor. Standard physical and engineering examples include both translational and rotational velocities and accelerations.

On the other hand, any geometrical object $v_i = v_i(t)$ that under the coordinate transformation (2.2) transforms as

$$\bar{v}_i = v_j \frac{\partial x^j}{\partial \bar{x}^i},$$

represents a *one-form* or *covector*, traditionally called a *covariant vector*, or, a first-order covariant tensor. Standard physical and engineering examples include both translational and rotational momenta, forces and torques.

2.1.1.4 Second-Order Tensors

Any geometrical object $t^{ik} = t^{ik}(t)$ that under the coordinate transformation (2.2) transforms as

$$\bar{t}^{ik} = t^{jl} \frac{\partial \bar{x}^i}{\partial x^j} \frac{\partial \bar{x}^k}{\partial x^l}, \quad (i, k = 1, \dots, n; j, l = 1, \dots, m),$$

represents a *second-order contravariant tensor*. It can be get as an *outer product* of two contravariant vectors, $t^{ik} = u^i v^k$.

Any geometrical object $t_{ik} = t_{ik}(t)$ that under the coordinate transformation (2.2) transforms as

$$\bar{t}_{ik} = t_{jl} \frac{\partial x^j}{\partial \bar{x}^i} \frac{\partial x^l}{\partial \bar{x}^k},$$

represents a *second-order covariant tensor*. It can be get as an outer product of two covariant vectors, $t_{ik} = u_i v_k$.

Any geometrical object $t_k^i = t_k^i(t)$ that under the coordinate transformation (2.2) transforms as

$$\bar{t}_k^i = t_l^j \frac{\partial \bar{x}^i}{\partial x^j} \frac{\partial x^l}{\partial \bar{x}^k},$$

represents a *second-order mixed tensor*. It can be get as an outer product of a covariant vector and a contravariant vector, $t_k^i = u^i v_k$.

Standard physical and engineering examples include:

- (1) The fundamental (material) covariant metric tensor $\mathbf{g} \equiv g_{ik}$, i.e., inertia matrix, given usually by the transformation from Cartesian coordinates

y^j to curvilinear coordinates x^i ,

$$g_{ik} = \frac{\partial y^j}{\partial x^i} \frac{\partial y^j}{\partial x^k}, \quad (\text{summing over } j). \quad (2.5)$$

It is used in the quadratic metric form ds^2 of the space in consideration (e.g., a certain physical or engineering configuration space)

$$ds^2 \equiv dy^j dy^j = g_{ik} dx^i dx^k,$$

where the first term on the r.h.s denotes the *Euclidean metrics*, while the second term is the *Riemannian metric* of the space, respectively.

(2) Its inverse $\mathbf{g}^{-1} \equiv g^{ik}$, given by

$$g^{ik} = (g_{ik})^{-1} = \frac{G_{ik}}{|g_{ik}|}, \quad G_{ik} \text{ is the cofactor of the matrix } (g_{ik});$$

(3) The *Kronecker-delta* symbol δ_k^i , given by

$$\delta_k^i = \begin{cases} 1 & \text{if } i = k \\ 0 & \text{if } i \neq k \end{cases},$$

used to denote the metric tensor in Cartesian orthogonal coordinates. δ_k^i is a discrete version of the *Dirac δ -function*. The *generalized Kronecker-delta symbol* δ_{lmn}^{ijk} (in 3D) is the product of *Ricci antisymmetric tensors* ε^{ijk} and ε_{lmn} ,

$$\delta_{lmn}^{ijk} = \varepsilon^{ijk} \varepsilon_{lmn} = \begin{cases} 0 & \text{if at least two indices are equal} \\ +1 & \text{if both } ijk \text{ and } lmn \text{ are either even or odd} \\ -1 & \text{if one of } ijk, lmn \text{ is even and the other is odd} \end{cases}.$$

For example, to derive components of the metric tensor $\mathbf{g} \equiv g_{ij}$ in standard spherical coordinates, we use the relations (2.3–2.4) between the spherical coordinates $x^i = \{\rho, \theta, \varphi\}$ and the Cartesian coordinates $y^k = \{x, y, z\}$, and the definition, $g_{ij} = \frac{\partial y^k}{\partial x^i} \frac{\partial y^k}{\partial x^j}$, to get the metric tensor (in matrix form)

$$(g_{ij}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & (\rho^1)^2 \cos^2 \varphi & 0 \\ 0 & 0 & (\rho^1)^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \rho^2 \cos^2 \varphi & 0 \\ 0 & 0 & \rho^2 \end{pmatrix}, \quad (2.6)$$

and the inverse metric tensor

$$(g^{ij}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{(x^1)^2 \cos^2 x^3} & 0 \\ 0 & 0 & \frac{1}{(x^1)^2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\rho^2 \cos^2 \varphi} & 0 \\ 0 & 0 & \frac{1}{\rho^2} \end{pmatrix}. \quad (2.7)$$

Given a tensor, we can derive other tensors by raising and lowering its indices, by their multiplication with covariant and contravariant metric tensors. In this way, the so-called *associated tensors* to the given tensor are formed. For example, v^i and v_i are associated tensors, related by

$$v_i = g_{ik} v^k \quad \text{and} \quad v^i = g^{ik} v_k.$$

Given two vectors, $\mathbf{u} \equiv u^i$ and $\mathbf{v} \equiv v^i$, their inner (dot, or scalar) product is given by

$$\mathbf{u} \cdot \mathbf{v} \equiv g_{ij} u^i v^j,$$

while their vector (cross) product (in 3D) is given by

$$\mathbf{u} \times \mathbf{v} \equiv \varepsilon_{ijk} u^j v^k.$$

2.1.1.5 Higher-Order Tensors

As a generalization of above tensors, consider a geometrical object $R_{kps}^i = R_{kps}^i(t)$ that under the coordinate transformation (2.2) transforms as

$$\bar{R}_{kps}^i = R_{lqt}^j \frac{\partial \bar{x}^i}{\partial x^j} \frac{\partial x^l}{\partial \bar{x}^k} \frac{\partial x^q}{\partial \bar{x}^p} \frac{\partial x^t}{\partial \bar{x}^s}, \quad (\text{all indices} = 1, \dots, n). \quad (2.8)$$

Clearly, $R_{kjl}^i = R_{kjl}^i(x, t)$ is a fourth order tensor, once contravariant and three times covariant, representing the central tensor in Riemannian geometry, called the *Riemann curvature tensor*. As all physical and engineering configuration spaces are Riemannian manifolds, they are all characterized by curvature tensors. In case $R_{kjl}^i = 0$, the corresponding Riemannian manifold reduces to the Euclidean space of the same dimension, in which $g_{ik} = \delta_k^i$.

If one contravariant and one covariant index of a tensor are set equal, the resulting sum is a tensor of rank two less than that of the original tensor. This process is called *tensor contraction*.

If to each point of a region in an n D space there corresponds a definite tensor, we say that a *tensor-field* has been defined. In particular, this is a *vector-field* or a *scalar-field* according as the tensor is of rank one or zero. It should be noted that a tensor or tensor-field is not just the set of its

components in one special coordinate system, but all the possible sets of components under any transformation of coordinates.

2.1.1.6 Tensor Symmetry

A tensor is called *symmetric* with respect to two indices of the same variance if its components remain unaltered upon interchange of the indices; e.g., $a_{ij} = a_{ji}$, or $a^{ij} = a^{ji}$. A tensor is called *skew-symmetric* (or, *antisymmetric*) with respect to two indices of the same variance if its components change sign upon interchange of the indices; e.g., $a_{ij} = -a_{ji}$, or $a^{ij} = -a^{ji}$. Regarding tensor symmetry, in the following we will prove several useful propositions.

(i) *Every second-order tensor can be expressed as the sum of two tensors, one of which is symmetric and the other is skew-symmetric.* For example, a second-order tensor a_{ij} , which is for $i, j = 1, \dots, n$ given by the $n \times n$ -matrix

$$a_{ij} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix},$$

can be rewritten as

$$\begin{aligned} a_{ij} &= \frac{1}{2}a_{ij} + \frac{1}{2}a_{ij} + \frac{1}{2}a_{ji} - \frac{1}{2}a_{ji}, \quad \text{that can be rearranged as} \\ &= \frac{1}{2}a_{ij} + \frac{1}{2}a_{ji} + \frac{1}{2}a_{ij} - \frac{1}{2}a_{ji}, \quad \text{which can be regrouped as} \\ &= \frac{1}{2}(a_{ij} + a_{ji}) + \frac{1}{2}(a_{ij} - a_{ji}), \quad \text{which can be written as} \\ &= a_{(ij)} + a_{[ij]}, \end{aligned}$$

where $a_{(ij)}$ denotes its symmetric part, while $a_{[ij]}$ denotes its skew-symmetric part, as required.

(ii) *Every quadratic form can be made symmetric.* For example, a quadratic form $a_{ij}x^i x^j$, that (for $i, j = 1, \dots, n$) expands as

$$\begin{aligned} a_{ij}x^i x^j &= a_{11}x^1 x^1 + a_{12}x^1 x^2 + \dots + a_{1n}x^1 x^n + \\ &\quad + a_{21}x^2 x^1 + a_{22}x^2 x^2 + \dots + a_{2n}x^2 x^n + \\ &\quad \dots \\ &\quad + a_{n1}x^n x^1 + a_{n2}x^n x^2 + \dots + a_{nn}x^n x^n, \end{aligned}$$

with a non-symmetric second-order tensor a_{ij} , can be made symmetric in

the following way.

$$a_{ij}x^i x^j = \frac{1}{2}a_{ij}x^i x^j + \frac{1}{2}a_{ij}x^i x^j.$$

If we swap indices in the second term, we get

$$= \frac{1}{2}a_{ij}x^i x^j + \frac{1}{2}a_{ji}x^j x^i, \quad \text{which is equal to}$$

$$= \frac{1}{2}(a_{ij} + a_{ji})x^i x^j.$$

If we now use a substitution,

$$\frac{1}{2}(a_{ij} + a_{ji}) \equiv b_{ij} = b_{ji}, \quad \text{we get}$$

$$a_{ij}x^i x^j = b_{ij}x^i x^j,$$

where a_{ij} is non-symmetric and b_{ij} is symmetric, as required.

(iii) *Every second-order tensor that is the sum $a^{ij} = u^i v^j + u^j v^i$, or, $a_{ij} = u_i v_j + u_j v_i$ is symmetric.* In both cases, if we swap the indices i and j , we get $a^{ji} = u^j v^i + u^i v^j$, (resp. $a_{ji} = u_j v_i + u_i v_j$), which implies that the tensor a^{ij} (resp. a_{ij}) is symmetric.

(iv) *Every second-order tensor that is the difference $b^{ij} = u^i v^j - u^j v^i$, or, $b_{ij} = u_i v_j - u_j v_i$ is skew-symmetric.* In both cases, if we swap the indices i and j , we get $b^{ji} = -(u^j v^i - u^i v^j)$, (resp. $b_{ji} = -(u_j v_i - u_i v_j)$), which implies that the tensor b^{ij} (resp. b_{ij}) is skew-symmetric.

2.1.2 Euclidean Tensors

2.1.2.1 Basis Vectors and the Metric Tensor in \mathbb{R}^n

The natural *Cartesian coordinate basis* in an n D Euclidean space \mathbb{R}^n is defined as a set of n D unit vectors e^i given by

$$e^1 = \{[1, 0, 0, \dots]^t, e^2 = \{0, 1, 0, \dots\}^t, e^3 = \{0, 0, 1, \dots\}^t, \dots, e^n = \{0, 0, \dots, 1\}^t\},$$

(where index t denotes transpose) while its dual basis e_i is given by:

$$e_1 = \{[1, 0, 0, \dots], e_2 = \{0, 1, 0, \dots\}, e_3 = \{0, 0, 1, \dots\}, \dots, e_n = \{0, 0, \dots, 1\}\},$$

(no transpose) where the definition of the dual basis is given by the Kronecker's δ -symbol, i.e., the $n \times n$ identity matrix:

$$e^i \cdot e_j = \delta_j^i = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix},$$

that is the metric tensor in Cartesian coordinates equals $\mathbf{g} = \delta_j^i$. In general, (i.e., curvilinear) coordinate system, the metric tensor $\mathbf{g} = g_{ij}$ is defined as the scalar product of the dual basis vectors, i.e., the $n \times n$ matrix:

$$g_{ij} = e_i \cdot e_j = \begin{bmatrix} g_{11} & g_{12} & g_{13} & \dots & g_{1n} \\ g_{21} & g_{22} & g_{23} & \dots & g_{2n} \\ g_{31} & g_{32} & g_{33} & \dots & g_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ g_{n1} & g_{n2} & g_{n3} & \dots & g_{nn} \end{bmatrix}.$$

2.1.2.2 Tensor Products in \mathbb{R}^n

Let \mathbf{u} and \mathbf{v} denote two vectors in \mathbb{R}^n , with their components given by

$$u^i = \mathbf{u} \cdot e^i, \quad \text{and} \quad v^j = \mathbf{v} \cdot e^j,$$

where $u = |\mathbf{u}|$ and $v = |\mathbf{v}|$ are their respective norms (or, lengths). Then their inner product (i.e., scalar, or dot product) $\mathbf{u} \cdot \mathbf{v}$ is a scalar invariant S , defined as

$$S = u^i \cdot v^j = g_{ij} u^i v^j.$$

Besides the dot product of two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, there is also their tensor product (i.e., generalized vector, or cross product), which is a second-order tensor

$$\mathbf{T} = \mathbf{u} \otimes \mathbf{v}, \text{ in components, } T^{ij} = u^i \otimes v^j.$$

In the natural basis e_i this tensor is expanded as

$$\mathbf{T} = T^{ij} e_i \otimes e_j,$$

while its components in the dual basis read:

$$T^{ij} = T(e^i, e^j),$$

where $T = |\mathbf{T}|$ is its norm. To get its components in curvilinear coordinates, we need first to substitute it in Cartesian basis:

$$T^{ij} = T^{mn}(e_m \otimes e_n)(e^i, e^j),$$

then to evaluate it on the slots:

$$T^{ij} = T^{mn} e_m \cdot e^i e_n \cdot e^j,$$

and finally to calculate the other index configurations by lowering indices, by means of the metric tensor:

$$T_j^i = g_{jm} T^{im}, \quad T_{ij} = g_{im} g_{jn} T^{mn}.$$

2.1.3 Covariant Differentiation

In this subsection, we need to consider some n D Riemannian manifold M (see section (3.10.1) below) with the metric form (i.e., line element) $ds^2 = g_{ik} dx^i dx^k$, as a configuration space for a certain physical or engineering system (e.g., robotic manipulator).

2.1.3.1 Christoffel's Symbols

Partial derivatives of the metric tensor g_{ik} (2.5) form themselves special symbols that do not transform as tensors (with respect to the coordinate transformation (2.2)), but nevertheless represent important quantities in tensor analysis. They are called *Christoffel symbols of the first kind*, defined by

$$\Gamma_{ijk} = \frac{1}{2}(\partial_{x^k} g_{ij} + \partial_{x^j} g_{ki} - \partial_{x^i} g_{jk}), \quad \left(\text{remember, } \partial_{x^i} \equiv \frac{\partial}{\partial x^i} \right)$$

and *Christoffel symbols of the second kind*, defined by

$$\Gamma_{ij}^k = g^{kl} \Gamma_{ijl}.$$

The Riemann curvature tensor R_{ijk}^l (2.8) of the manifold M , can be expressed in terms of the later as

$$R_{ijk}^l = \partial_{x^j} \Gamma_{ik}^l - \partial_{x^k} \Gamma_{ij}^l + \Gamma_{rj}^l \Gamma_{ik}^r - \Gamma_{rk}^l \Gamma_{ij}^r.$$

For example, in 3D spherical coordinates, $x^i = \{\rho, \theta, \varphi\}$, with the metric tensor and its inverse given by (2.6, 2.7), it can be shown that the only

nonzero Christoffel's symbols are:

$$\begin{aligned}\Gamma_{12}^2 &= \Gamma_{21}^2 = \Gamma_{13}^3 = \Gamma_{31}^3 = \frac{1}{\rho}, & \Gamma_{23}^3 &= \Gamma_{32}^2 = -\tan \theta, \\ \Gamma_{22}^1 &= -\rho, & \Gamma_{33}^1 &= -\rho \cos^2 \theta, & \Gamma_{33}^2 &= \sin \theta \cos \theta.\end{aligned}\quad (2.9)$$

2.1.3.2 Geodesics

From the Riemannian metric form $ds^2 = g_{ik}dx^i dx^k$ it follows that the distance between two points t_1 and t_2 on a curve $x^i = x^i(t)$ in M is given by

$$s = \int_{t_1}^{t_2} \sqrt{g_{ik} \dot{x}^i \dot{x}^k} dt.$$

That curve $x^i = x^i(t)$ in M which makes the distance s a minimum is called a *geodesic* of the space M (e.g., in a sphere, the geodesics are arcs of great circles). Using the calculus of variations, the geodesics are found from the differential *geodesic equation*,

$$\ddot{x}^i + \Gamma_{jk}^i \dot{x}^j \dot{x}^k = 0, \quad (2.10)$$

where overdot means derivative upon the line parameter s .

For example, in 3D spherical coordinates $x^i = \{\rho, \theta, \varphi\}$, using (2.9), geodesic equation (2.10) becomes a system of three scalar ODEs,

$$\begin{aligned}\ddot{\rho} - \rho \dot{\theta}^2 - \rho \cos^2 \theta \dot{\varphi}^2 &= 0, & \ddot{\theta} + \frac{2}{\rho} \dot{\rho} \dot{\theta} + \sin \theta \cos \theta \dot{\varphi}^2 &= 0, \\ \ddot{\varphi} + \frac{2}{\rho} \dot{\rho} \dot{\varphi} - 2 \tan \theta \dot{\theta} \dot{\varphi} &= 0.\end{aligned}\quad (2.11)$$

2.1.3.3 Covariant Derivative

Ordinary total and partial derivatives of vectors (covectors) *do not transform as vectors* (covectors) with respect to the coordinate transformation (2.2). For example, let y^k be Cartesian coordinates and x^i be general curvilinear coordinates of a dynamical system (with $i, k = 1, \dots, n$). We have: $x^i(t) = x^i[y^k(t)]$, which implies that

$$\frac{dx^i}{dt} = \frac{\partial x^i}{\partial y^k} \frac{dy^k}{dt}, \quad \text{or equivalently,} \quad \dot{x}^i = \frac{\partial x^i}{\partial y^k} \dot{y}^k,$$

that is a transformation law for the contravariant vector, which means that the velocity $v^i \equiv \dot{x}^i \equiv \frac{dx^i}{dt}$ is a proper contravariant vector. However, if we

perform another time differentiation, we get

$$\frac{d^2 x^i}{dt^2} = \frac{\partial x^i}{\partial y^k} \frac{d^2 y^k}{dt^2} + \frac{\partial^2 x^i}{\partial y^k \partial y^m} \frac{dy^k}{dt} \frac{dy^m}{dt},$$

which means that $\frac{d^2 x^i}{dt^2}$ is not a proper vector.

$\frac{d^2 x^i}{dt^2}$ is an acceleration vector only in a special case when x^i are another Cartesian coordinates; then $\frac{\partial^2 x^i}{\partial y^k \partial y^m} = 0$, and therefore the original coordinate transformation is linear, $x^i = a_k^i y^k + b^i$ (where a_k^i and b^i are constant).

Therefore, $\frac{d^2 x^i}{dt^2}$ represents an acceleration vector only in terms of Newtonian mechanics in a Euclidean space \mathbb{R}^n , while it is not a proper acceleration vector in terms of Lagrangian or Hamiltonian mechanics in general curvilinear coordinates on a smooth manifold M^n . And we know that Newtonian mechanics in \mathbb{R}^n is sufficient only for fairly simple mechanical systems.

The above is true for any tensors. So we need to find another derivative operator to be able to preserve their tensor character. The solution to this problem is called the *covariant derivative*.

The covariant derivative $v_{;k}^i$ of a contravariant vector v^i is defined as

$$v_{;k}^i = \partial_{x^k} v^i + \Gamma_{jk}^i v^j.$$

Similarly, the covariant derivative $v_{i;k}$ of a covariant vector v_i is defined as

$$v_{i;k} = \partial_{x^k} v_i - \Gamma_{ik}^j v_j.$$

Generalization for the higher order tensors is straightforward; e.g., the covariant derivative $t_{kl;q}^j$ of the third order tensor t_{kl}^j is given by

$$t_{kl;q}^j = \partial_{x^q} t_{kl}^j + \Gamma_{qs}^j t_{kl}^s - \Gamma_{kq}^s t_{sl}^j - \Gamma_{lq}^s t_{ks}^j.$$

The covariant derivative is the most important tensor operator in general relativity (its zero defines *parallel transport*) as well as the basis for defining other differential operators in mechanics and physics.

2.1.3.4 Covariant Form of Differential Operators

Here we give the covariant form of classical vector differential operators: gradient, divergence, curl and Laplacian.

Gradient. If $\varphi = \varphi(x^i, t)$ is a scalar field, the gradient one-form $\text{grad}(\varphi)$ is defined by

$$\text{grad}(\varphi) = \nabla \varphi = \varphi_{;i} = \partial_{x^i} \varphi.$$

Divergence. The divergence $\text{div}(v^i)$ of a vector-field $v^i = v^i(x^i, t)$ is defined by contraction of its covariant derivative with respect to the coordinates $x^i = x^i(t)$, i.e., the contraction of $v^i_{;k}$, namely

$$\text{div}(v^i) = v^i_{;i} = \frac{1}{\sqrt{g}} \partial_{x^i} (\sqrt{g} v^i).$$

Curl. The curl $\text{curl}(\theta_i)$ of a one-form $\theta_i = \theta_i(x^i, t)$ is a second-order covariant tensor defined as

$$\text{curl}(\theta_i) = \theta_{i;k} - \theta_{k;i} = \partial_{x^k} \theta_i - \partial_{x^i} \theta_k.$$

Laplacian. The Laplacian $\Delta\varphi$ of a scalar invariant $\varphi = \varphi(x^i, t)$ is the divergence of $\text{grad}(\varphi)$, or

$$\Delta\varphi = \nabla^2\varphi = \text{div}(\text{grad}(\varphi)) = \text{div}(\varphi_{;i}) = \frac{1}{\sqrt{g}} \partial_{x^i} (\sqrt{g} g^{ik} \partial_{x^k} \varphi).$$

2.1.3.5 Absolute Derivative

The *absolute derivative* (or *intrinsic*, or *Bianchi's derivative*) of a contravariant vector v^i along a curve $x^k = x^k(t)$ is denoted by $\dot{v}^i \equiv Dv^i/dt$ and defined as the inner product of the covariant derivative of v^i and $\dot{x}^k \equiv dx^k/dt$, i.e., $v^i_{;k} \dot{x}^k$, and is given by

$$\dot{v}^i = v^i + \Gamma^i_{jk} v^j \dot{x}^k.$$

Similarly, the absolute derivative \dot{v}_i of a covariant vector v_i is defined as

$$\dot{v}_i = v_i - \Gamma^j_{ik} v_j \dot{x}^k.$$

Generalization for the higher order tensors is straightforward; e.g., the absolute derivative \dot{t}^j_{kl} of the third order tensor t^j_{kl} is given by

$$\dot{t}^j_{kl} = t^j_{kl} + \Gamma^j_{qs} t^s_{kl} \dot{x}^q - \Gamma^s_{kq} t^j_{sl} \dot{x}^q - \Gamma^s_{lq} t^j_{ks} \dot{x}^q.$$

The absolute derivative is the most important differential operator in physics and engineering, as it is the basis for the *covariant form* of both Lagrangian and Hamiltonian equations of motion of many physical and engineering systems.

2.1.3.6 3D Curve Geometry: Frenet–Serret Formulae

Given three unit vectors: *tangent* τ^i , *principal normal* β^i , and *binormal* ν^i , as well as two scalar invariants: curvature K and torsion T , of a curve $\gamma(s) = \gamma[x^i(s)]$, the so-called *Frenet–Serret formulae* are valid¹

$$\begin{aligned}\dot{\tau}^i &\equiv \dot{\tau}^i + \Gamma_{jk}^i \tau^j \dot{x}^k = K\beta^i, \\ \dot{\beta}^i &\equiv \dot{\beta}^i + \Gamma_{jk}^i \beta^j \dot{x}^k = -(K\tau^i + T\nu^i), \\ \dot{\nu}^i &\equiv \dot{\nu}^i + \Gamma_{jk}^i \nu^j \dot{x}^k = T\beta^i.\end{aligned}$$

2.1.3.7 Mechanical Acceleration and Force

In modern analytical mechanics, the two fundamental notions of *acceleration* and *force* in general curvilinear coordinates are substantially different from the corresponding terms in Cartesian coordinates as commonly used in engineering mechanics. Namely, the acceleration vector *is not* an ordinary time derivative of the velocity vector; ‘even worse’, the force, which is a paradigm of a vector in statics and engineering vector mechanics, *is not* a vector at all. Proper mathematical definition of the acceleration vector is the absolute time derivative of the velocity vector, while the force is a differential one-form.

To give a brief look at these ‘weird mathematical beasts’, consider a material dynamical system described by n curvilinear coordinates $x^i = x^i(t)$. First, recall from section 2.1.3.3 above, that an ordinary time derivative of the velocity vector $v^i(t) = \dot{x}^i(t)$ *does not transform as a vector* with respect to the general coordinate transformation (2.2). Therefore, $a^i \neq \dot{v}^i$. So, we need to use its absolute time derivative to define the acceleration vector (with $i, j, k = 1, \dots, n$),

$$a^i = \dot{v}^i \equiv \frac{Dv^i}{dt} = v_{;k}^i \dot{x}^k \equiv \dot{v}^i + \Gamma_{jk}^i v^j \dot{x}^k \equiv \ddot{x}^i + \Gamma_{jk}^i \dot{x}^j \dot{x}^k, \quad (2.12)$$

which is equivalent to the l.h.s of the geodesic equation (2.10). Only in the particular case of Cartesian coordinates, the general acceleration vector (2.12) reduces to the familiar engineering form of the Euclidean acceleration vector², $\mathbf{a} = \dot{\mathbf{v}}$.

¹In this paragraph, the overdot denotes the total derivative with respect to the line parameter s (instead of time t).

²Any Euclidean space can be defined as a set of Cartesian coordinates, while any Riemannian manifold can be defined as a set of curvilinear coordinates. Christoffel’s symbols Γ_{jk}^i vanish in Euclidean spaces defined by Cartesian coordinates; however, they are nonzero in Riemannian manifolds defined by curvilinear coordinates.

For example, in standard spherical coordinates $x^i = \{\rho, \theta, \varphi\}$, we have the components of the acceleration vector given by (2.11), if we now reinterpret overdot as the time derivative,

$$\begin{aligned} a^\rho &= \ddot{\rho} - \rho \dot{\theta}^2 - \rho \cos^2 \theta \dot{\varphi}^2, & a^\theta &= \ddot{\theta} + \frac{2}{\rho} \dot{\rho} \dot{\varphi} + \sin \theta \cos \theta \dot{\varphi}^2, \\ a^\varphi &= \ddot{\varphi} + \frac{2}{\rho} \dot{\rho} \dot{\varphi} - 2 \tan \theta \dot{\theta} \dot{\varphi}. \end{aligned}$$

Now, using (2.12), the *Newton's fundamental equation of motion*, that is the basis of all science, $\mathbf{F} = m \mathbf{a}$, gets the following tensorial form

$$F^i = m a^i = m \dot{v}^i = m(v^i_{;k} \dot{x}^k) \equiv m(\dot{v}^i + \Gamma^i_{jk} v^j \dot{x}^k) = m(\ddot{x}^i + \Gamma^i_{jk} \dot{x}^j \dot{x}^k), \quad (2.13)$$

which defines Newtonian force as a contravariant vector.

However, modern Hamiltonian dynamics reminds us that: (i) Newton's own force definition was not really $\mathbf{F} = m \mathbf{a}$, but rather $\mathbf{F} = \dot{\mathbf{p}}$, where \mathbf{p} is the system's momentum, and (ii) the momentum \mathbf{p} is not really a vector, but rather a dual quantity, a differential one-form³. Consequently, the force, as its time derivative, is also a one-form (see Figure 2.1; also, compare with Figure Figure 5.2 above). This new force definition includes the precise definition of the mass distribution within the system, by means of its Riemannian metric tensor g_{ij} . Thus, (2.13) has to be modified as

$$F_i = m g_{ij} a^j \equiv m g_{ij} (\dot{v}^j + \Gamma^j_{ik} v^i \dot{x}^k) = m g_{ij} (\ddot{x}^j + \Gamma^j_{ik} \dot{x}^i \dot{x}^k), \quad (2.14)$$

where the quantity $m g_{ij}$ is called the *material metric tensor*, or *inertia matrix*. Equation (2.14) generalizes the notion of the Newtonian force \mathbf{F} , from Euclidean space \mathbb{R}^n to the Riemannian manifold M .

2.1.4 Application: Covariant Mechanics

Recall that a material system is regarded from the dynamical standpoint as a collection of particles which are subject to interconnections and constraints of various kinds (e.g., a rigid body is regarded as a number of particles rigidly connected together so as to remain at invariable distances from each other). The *number of independent coordinates which determine the configuration of a dynamical system completely* is called the *number of degrees of freedom* (DOF) of the system. In other words, this number, n ,

³For example, in Dirac's $\langle bra|ket \rangle$ formalism, *kets* are vectors, while *bras* are one-forms; in matrix notation, columns are vectors, while rows are one-forms.

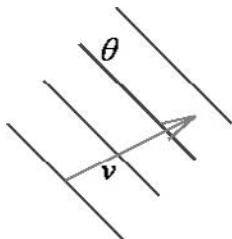


Fig. 2.1 A one-form θ (which is a family of parallel (hyper)surfaces, the so-called *Grassmann planes*) pierced by the vector v to give a scalar product $\theta(v) \equiv \langle \theta, v \rangle = 2.6$ (see text for explanation).

is the *dimension of the system's configuration manifold*. This viewpoint is the core of our applied differential geometry.

For simplicity, let us suppose that we have a dynamical system with three DOF (e.g., a particle of mass M , or a rigid body of mass M with one point fixed); generalization to n DOF, with N included masses M_α , is straightforward. The configuration of our system at any time is then given by three coordinates $\{q^1, q^2, q^3\}$. As the coordinates change in value the dynamical system changes its configuration. Obviously, there is an infinite number of sets of independent coordinates which will determine the configuration of a dynamical system, but since the position of the system is completely given by any one set, these sets of coordinates must be functionally related. Hence, if \bar{q}^i is any other set of coordinates, these quantities must be connected with q^i by formulae of the type

$$\bar{q}^i = \bar{q}^i(q^i), \quad (i = 1, \dots, n(= 3)). \quad (2.15)$$

Relations (2.15) are the equations of transformation from one set of dynamical coordinates to another and, in a standard tensorial way (see [Misner *et al.* (1973)]), we can define tensors relative to this coordinate transformation. The generalized coordinates q^i , ($i = 1, \dots, n$) constitute the system's *configuration manifold*.

In particular, in our ordinary Euclidean 3-dimensional (3D) space \mathbb{R}^3 , the ordinary Cartesian axes are $x^i = \{x, y, z\}$, and the *system's center of mass* (COM) is given by

$$C^i = \frac{M_\alpha x_\alpha^i}{\sum_{\alpha=1}^N M_\alpha},$$

where Greek subscript α labels the masses included in the system. If we have a continuous distribution of matter $V = V(M)$ rather than the dis-

crete system of masses M_α , all the α -sums should be replaced by volume integrals, the element of mass dM taking the place of M_α ,

$$\sum_{\alpha=1}^N M_\alpha \Rightarrow \iiint_V dM.$$

An important quantity related to the system's COM is the double symmetric contravariant tensor

$$I^{ij} = M_\alpha x_\alpha^i x_\alpha^j, \quad (2.16)$$

called the *inertia tensor*, calculated relative to the origin O of the Cartesian axes $x_\alpha^i = \{x_\alpha, y_\alpha, z_\alpha\}$. If we are given a straight line through O , defined by its unit vector λ^i , and perpendiculars p_α are drawn from the different particles on the line λ^i , the quantity

$$I(\lambda^i) = M_\alpha p_\alpha^2$$

is called the *moment of inertia* around λ^i . The moment of inertia $I(\lambda^i)$ can be expressed through inertia tensor (2.16) as

$$I(\lambda^i) = (I g_{ij} - I_{ij}) \lambda^i \lambda^j,$$

where g_{ij} is the system's *Euclidean 3D metric tensor* (as defined above), $I = g_{ij} I^{ij}$, and $I_{ij} = g_{rm} g_{sn} I^{mn}$ is the *covariant inertia tensor*. If we now consider the quadric Q whose equation is

$$(I g_{ij} - I_{ij}) x^i x^j = 1, \quad (2.17)$$

we find that the moment of inertia around λ^i is $1/R$, where R is the radius vector of Q in the direction of λ^i . The quadric Q defined by relation (2.17) is called the *ellipsoid of inertia* at the origin O . It has always three principal axes, which are called the *principal axes of inertia* at O , and the planes containing them in pairs are called the *principal planes of inertia* at O . The principal axes of inertia are given by the equations

$$(I g_{ij} - I_{ij}) \lambda^j = \theta \lambda^i,$$

where θ is a root of the determinant equation

$$|(I - \theta) g_{ij} - I_{ij}| = 0.$$

More generally, if we suppose that the points of our dynamical system are referred to rectilinear Cartesian axes x^i in a Euclidean n -dimensional

(nD) space \mathbb{R}^n , then when we are given the time and a set of *generalized coordinates* q^i we are also given all the points x^i of the dynamical system, as the system is determined uniquely. Consequently, the x^i are functions of q^i and possibly also of the time, that is,

$$x^i = x^i(q^i, t).$$

If we restrict ourselves to the *autonomous dynamical systems* in which these equations do not involve t , i.e.,

$$x^i = x^i(q^i), \quad (2.18)$$

then differentiating (2.18) with respect to the time t gives

$$\dot{x}^i = \frac{\partial x^i}{\partial q^j} \dot{q}^j. \quad (2.19)$$

The quantities \dot{q}^i , which form a vector with reference to coordinate transformations (2.15), we shall call the *generalized velocity vector*. We see from (2.19) that when the generalized velocity vector is given we know the velocity of each point of our system. Further, this gives us the system's *kinetic energy*,

$$E_{kin} = \frac{1}{2} M_\alpha g_{mn} \dot{x}_\alpha^m \dot{x}_\alpha^n = \frac{1}{2} M_\alpha g_{mn} \frac{\partial x_\alpha^m}{\partial q^i} \frac{\partial x_\alpha^n}{\partial q^j} \dot{q}^i \dot{q}^j. \quad (2.20)$$

Now, if we use the Euclidean metric tensor g_{ij} to define the *material metric tensor* G_{ij} , including the distribution of all the masses M_α of our system, as

$$G_{ij} = M_\alpha g_{mn} \frac{\partial x_\alpha^m}{\partial q^i} \frac{\partial x_\alpha^n}{\partial q^j}, \quad (2.21)$$

the kinetic energy (2.20) becomes a *homogenous quadratic form* in the generalized system's velocities \dot{q}^i ,

$$E_{kin} = \frac{1}{2} G_{ij} \dot{q}^i \dot{q}^j. \quad (2.22)$$

From the transformation relation (2.21) we see that the material metric tensor G_{ij} is symmetric in i and j . Also, since E_{kin} is an invariant for all transformations of generalized coordinates, from (2.22) we conclude that G_{ij} is a double symmetric tensor. Clearly, this is the central quantity in classical tensor system dynamics. We will see later, that G_{ij} defines the Riemannian geometry of the system dynamics. For simplicity reasons,

G_{ij} is often denoted by purely geometrical symbol g_{ij} , either assuming or neglecting the material properties of the system.

Now, let us find the equations of motion of our system. According to the *D'Alembert's Principle of virtual displacements*, the equations of motion in Cartesian coordinates x^i in \mathbb{R}^n are embodied in the single tensor equation

$$g_{mn}(M_\alpha \ddot{x}_\alpha^m - X_\alpha^m) \delta x_\alpha^n = 0, \quad (2.23)$$

where X_α^i is the total force vector acting on the particle M_α , while δx_α^i is the associated *virtual displacement* vector, so that the product $g_{ij} X_\alpha^i \delta x_\alpha^j$ is the *virtual work* of the system, and we can neglect in X_α^i all the internal or external forces which do not work in the displacement δx_α^i . If we give the system a small displacement compatible to with the constraints of the system, we see that this displacement may be effected by giving increments δq^i to the generalized coordinates q^i of the system, and these are related to the δx^i in accordance with the transformation formulae $\delta x_\alpha^i = \frac{\partial x_\alpha^i}{\partial q^j} \delta q^j$.

Furthermore, in this displacement the internal forces due to the constraints of the system will do no work, since these constraints are preserved, and consequently only the external forces will appear in (2.23), so it becomes

$$g_{mn} \left[M_\alpha \frac{d}{dt} \left(\frac{\partial x_\alpha^m}{\partial q^j} \dot{q}^j \right) \frac{\partial x_\alpha^n}{\partial q^i} - X_\alpha^m \frac{\partial x_\alpha^n}{\partial q^i} \right] \delta q^i = 0. \quad (2.24)$$

Now, using (2.20–2.22), we derive

$$M_\alpha g_{mn} \frac{d}{dt} \left(\frac{\partial x_\alpha^m}{\partial q^j} \dot{q}^j \right) \frac{\partial x_\alpha^n}{\partial q^i} = \frac{d}{dt} (G_{ij} \dot{q}^j) - \frac{1}{2} \frac{\partial G_{st}}{\partial q^i} \dot{q}^j \dot{q}^k = \frac{d}{dt} \left(\frac{\partial E_{kin}}{\partial \dot{q}^i} \right) - \frac{\partial E_{kin}}{\partial q^i}.$$

Also, if we put

$$\begin{aligned} F_i &= g_{mn} X_\alpha^m \frac{\partial x_\alpha^n}{\partial q^i}, & \text{we get} \\ F_i \delta q^i &= g_{mn} X_\alpha^m \delta x_\alpha^n = \delta W, \end{aligned} \quad (2.25)$$

where δW is the virtual work done by the external forces in the small displacement δq^i , which shows that F_i is the covariant vector, called the *generalized force vector*. Now (2.24) takes the form

$$\left[\frac{d}{dt} \left(\frac{\partial E_{kin}}{\partial \dot{q}^i} \right) - \frac{\partial E_{kin}}{\partial q^i} - F_i \right] \delta q^i = 0.$$

Since the coordinates q^i are independent this equation is true for all variations δq^i and we get as a final result the *covariant Lagrangian equations of motion*,

$$\frac{d}{dt} \left(\frac{\partial E_{kin}}{\partial \dot{q}^i} \right) - \frac{\partial E_{kin}}{\partial q^i} = F_i.$$

If the force system is conservative and E_{pot} is the system's *potential energy* given by

$$F_i = - \frac{\partial E_{pot}}{\partial q^i},$$

then, using (2.25) the Lagrangian equations take the standard form

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}^i} \right) = \frac{\partial L}{\partial q^i}, \quad (2.26)$$

where the *Lagrangian function* $L = L(q, \dot{q})$ of the system is given by $L = E_{kin} - E_{pot}$ (since E_{pot} does not contain \dot{q}^i).

Now, the kinetic energy E_{kin} of the system, given by quadratic form (2.22), is always positive except when \dot{q}^i is zero in which case E_{kin} vanishes. In other words, the quadratic form (2.22) is positive definite. Consequently, we can always find the *line (or arc) element*, defined by

$$ds^2 = G_{ij} dq^i dq^j. \quad (2.27)$$

A manifold in which ds^2 is given by relation of the type of (2.27), geometrically with g_{ij} instead of G_{ij} , is called a *Riemannian manifold*.

2.1.4.1 Riemannian Curvature Tensor

Every Riemannian manifold is characterized by the *Riemann curvature tensor*. In physical literature (see, e.g., [Misner *et al.* (1973)]) it is usually introduced through the *Jacobi equation of geodesic deviation*, showing the acceleration of the relative separation of nearby geodesics (the shortest, straight lines on the manifold). For simplicity, consider a sphere of radius a in \mathbb{R}^3 . Here, Jacobi equation is pretty simple,

$$\frac{d^2 \xi}{ds^2} + R\xi = 0,$$

where ξ is the *geodesic separation vector* (the so-called Jacobi vector-field), s denotes the geodesic arc parameter given by (2.27) and $R = 1/a^2$ is the *Gaussian curvature* of the surface.

In case of a higher-dimensional manifold M , the situation is naturally more complex, but the main structure of the Jacobi equation remains similar,

$$\frac{D^2\xi}{ds^2} + R(u, \xi, u) = 0,$$

where D denotes the covariant derivative and $R(u, \xi, u)$ is the curvature tensor, a three-slot linear machine. In components defined in a local coordinate chart (x^i) on M , this equation reads

$$\frac{D^2\xi^i}{ds^2} + R^i_{jkl} \frac{dx^j}{ds} \xi^k \frac{dx^l}{ds} = 0,$$

where R^i_{jkl} are the components of the *Riemann curvature tensor*.

2.1.4.2 Exterior Differential Forms

Recall that *exterior differential forms* are a special kind of antisymmetrical covariant tensors (see, e.g., [De Rham (1984); Flanders (1963)]). Such tensor-fields arise in many applications in physics, engineering, and differential geometry. The reason for this is the fact that the classical vector operations of **grad**, **div**, and **curl** as well as the theorems of Green, Gauss, and Stokes can all be expressed concisely in terms of differential forms and the main operator acting on them, the exterior derivative d . Differential forms inherit all geometrical properties of the general tensor calculus and add to it their own powerful geometrical, algebraic and topological machinery (see Figures 2.2 and 2.3). Differential p -forms formally occur as *integrands* under ordinary integral signs in \mathbb{R}^3 :

- a *line integral* $\int P dx + Q dy + R dz$ has as its integrand the *one-form* $\omega = P dx + Q dy + R dz$;
- a *surface integral* $\iint A dydz + B dzdx + C dxdy$ has as its integrand the *two-form* $\alpha = A dydz + B dzdx + C dxdy$;
- a *volume integral* $\iiint K dxdydz$ has as its integrand the *three-form* $\lambda = K dxdydz$.

By means of an *exterior derivative* d , a *derivation* that transforms p -forms into $(p+1)$ -forms, these geometrical objects generalize ordinary vector differential operators in \mathbb{R}^3 :

- a *scalar function* $f = f(x)$ is a zero-form;

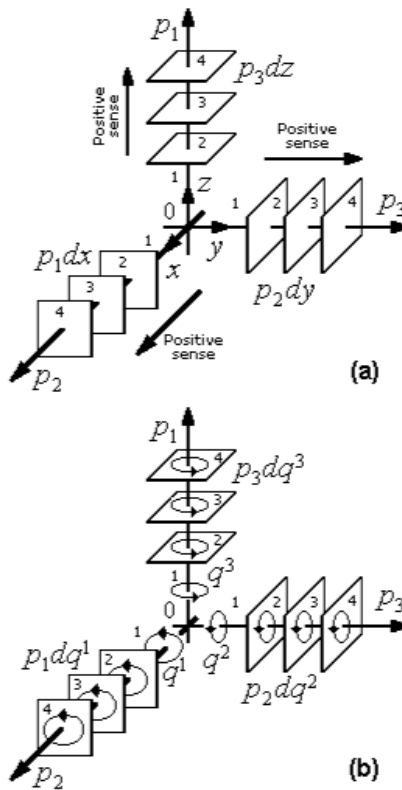


Fig. 2.2 Basis vectors and 1-forms in Euclidean \mathbb{R}^3 -space: (a) Translational case; and (b) Rotational case.

- its *gradient* df , is a one-form⁴

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz;$$

- a *curl* $d\omega$, of a one-form ω above, is a two-form

$$d\omega = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z} \right) dydz + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x} \right) dzdx + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dxdy;$$

⁴We use the same symbol, d , to denote both ordinary and exterior derivation, in order to avoid extensive use of the boldface symbols. It is clear from the context which derivative (differential) is in place: exterior derivative operates only on differential forms, while the ordinary differential operates mostly on coordinates.

- a *divergence* $d\alpha$, of the two-form α above, is a three-form

$$d\alpha = \left(\frac{\partial A}{\partial x} + \frac{\partial B}{\partial y} + \frac{\partial C}{\partial z} \right) dx dy dz.$$

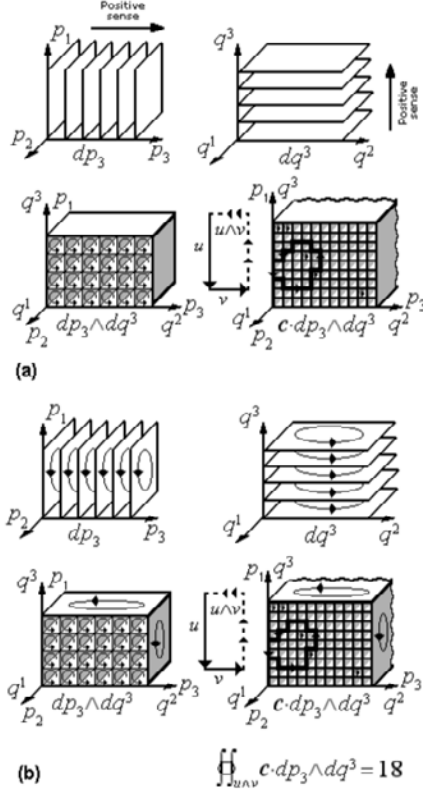


Fig. 2.3 Fundamental two-form and its flux in \mathbb{R}^3 : (a) Translational case; (b) Rotational case. In both cases the flux through the plane $u \wedge v$ is defined as $\int \int_{u \wedge v} c \cdot dp_i dq^i$ and measured by the number of tubes crossed by the circulation oriented by $u \wedge v$.

Now, although visually intuitive, our Euclidean 3D space \mathbb{R}^3 is not sufficient for thorough physical or engineering analysis. The fundamental concept of a *smooth manifold*, locally topologically equivalent to the Euclidean n D space \mathbb{R}^n , is required (with or without Riemannian metric tensor defined on it). In general, a proper definition of exterior derivative d for a p -form β on a smooth manifold M , includes the *Poincaré lemma*:

$d(d\beta) = 0$, and validates the general integral *Stokes formula*

$$\int_{\partial M} \beta = \int_M d\beta,$$

where M is a p -dimensional *manifold with a boundary* and ∂M is its $(p - 1)$ -dimensional *boundary*, while the integrals have appropriate dimensions.

A p -form β is called *closed* if its exterior derivative is equal to zero,

$$d\beta = 0.$$

From this condition one can see that the closed form (the *kernel* of the exterior derivative operator d) is conserved quantity. Therefore, closed p -forms possess certain invariant properties, physically corresponding to the conservation laws.

A p -form β that is an exterior derivative of some $(p - 1)$ -form α ,

$$\beta = d\alpha,$$

is called *exact* (the *image* of the exterior derivative operator d). By *Poincaré lemma*, exact forms prove to be closed automatically,

$$d\beta = d(d\alpha) = 0.$$

Similarly to the components of a 3D vector v defined above, a one-form θ defined on an n D manifold M can also be expressed in components, using the coordinate basis $\{dx^i\}$ along the local n D coordinate chart $\{x^i\} \in M$, as

$$\theta = \theta_i dx^i.$$

Now, the components of the exterior derivative of θ are equal to the components of its *commutator* defined on M by

$$d\theta = \omega_{ij} dx^i dx^j,$$

where the components of the *form commutator* ω_{ij} are given by

$$\omega_{ij} = \left(\frac{\partial \theta_j}{\partial x^i} - \frac{\partial \theta_i}{\partial x^j} \right).$$

The space of all smooth p -forms on a smooth manifold M is denoted by $\Omega^p(M)$. The *wedge*, or *exterior product* of two differential forms, a p -form

$\alpha \in \Omega^p(M)$ and a q -form $\beta \in \Omega^q(M)$ is a $(p+q)$ -form $\alpha \wedge \beta$. For example, if $\theta = a_i dx^i$, and $\eta = b_j dx^j$, their wedge product $\theta \wedge \eta$ is given by

$$\theta \wedge \eta = a_i b_j dx^i dx^j,$$

so that the coefficients $a_i b_j$ of $\theta \wedge \eta$ are again smooth functions, being polynomials in the coefficients a_i of θ and b_j of η . The exterior product \wedge is related to the exterior derivative $d : \Omega^p(M) \rightarrow \Omega^{p+1}(M)$, by

$$d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^p \alpha \wedge d\beta.$$

Another important linear operator is the *Hodge star* $*$: $\Omega^p(M) \rightarrow \Omega^{n-p}(M)$, where n is the dimension of the manifold M . This operator depends on the inner product (i.e., Riemannian metric) on M and also depends on the orientation (reversing orientation will change the sign). For any p -forms α and β ,

$$**\alpha = (-1)^{p(n-p)}\alpha, \quad \text{and} \quad \alpha \wedge *\beta = \beta \wedge *\alpha.$$

Hodge star is generally used to define *dual* $(n-p)$ -forms on n D smooth manifolds.

For example, in \mathbb{R}^3 with the ordinary Euclidean metric, if f and g are functions then (compare with the 3D forms of gradient, curl and divergence defined above)

$$\begin{aligned} df &= \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz, \\ *df &= \frac{\partial f}{\partial x} dydz + \frac{\partial f}{\partial y} dzdx + \frac{\partial f}{\partial z} dxdy, \\ df \wedge *dg &= \left(\frac{\partial f}{\partial x} \frac{\partial g}{\partial x} + \frac{\partial f}{\partial y} \frac{\partial g}{\partial y} + \frac{\partial f}{\partial z} \frac{\partial g}{\partial z} \right) dxdydz = \Delta f dxdydz, \end{aligned}$$

where Δf is the *Laplacian* on \mathbb{R}^3 . Therefore the three-form $df \wedge *dg$ is the Laplacian multiplied by the volume element, which is valid, more generally, in any local orthogonal coordinate system in any smooth domain $U \in \mathbb{R}^3$.

The subspace of all closed p -forms on M we will denote by $Z^p(M) \subset \Omega^p(M)$, and the sub-subspace of all exact p -forms on M we will denote by $B^p(M) \subset Z^p(M)$. Now, the quotient space

$$H^p(M) = \frac{Z^p(M)}{B^p(M)} = \frac{\text{Ker}(d : \Omega^p(M) \rightarrow \Omega^{p+1}(M))}{\text{Im}(d : \Omega^{p-1}(M) \rightarrow \Omega^p(M))}$$

is called the p th *de Rham cohomology group* (or vector space) of a manifold M . Two p -forms α and β on M are equivalent, or belong to the same *cohomology class* $[\alpha] \in H^p(M)$, if their difference equals $\alpha - \beta = d\theta$, where θ is a $(p - 1)$ -form on M .

2.1.4.3 The Covariant Force Law

Objective of this final tensor section is to generalize the fundamental Newtonian 3D equation, $\mathbf{F} = m\mathbf{a}$, for a generic robotic/biodynamical system, consisting of a number of flexibly-coupled rigid segments (see Figures 3.7–3.8 above), and thus to formulate the *covariant force law*.

To be able to apply the covariant formalism, we need to start with the suitable coordinate transformation (2.2), in this case as a relation between the 6 *external* $SE(3)$ rigid-body coordinates, $y^e = y^e(t)$ ($e = 1, \dots, 6$), and $2n$ *internal joint coordinates*, $x^i = x^i(t)$ ($i = 1, \dots, 2n$) (n angles, forming the constrained n -torus T^n , plus n very restricted translational coordinates, forming the hypercube $I^n \subset \mathbb{R}^n$). Once we have these two sets of coordinates, external- y^e and internal- x^i , we can perform the general functional transformation (2.2) between them,

$$x^i = x^i(y^e). \quad (2.28)$$

Now, although the coordinate transformation (2.28) is nonlinear and even unknown at this stage, there is something known and simple about it: the corresponding transformation of differentials is *linear and homogenous*,

$$dx^i = \frac{\partial x^i}{\partial y^e} dy^e,$$

which implies the *linear and homogenous transformation of velocities*,

$$\dot{x}^i = \frac{\partial x^i}{\partial y^e} \dot{y}^e. \quad (2.29)$$

Our *internal velocity vector-field* is defined by the set of ODEs (2.29), at each *representative point* $x^i = x^i(t)$ of the system's configuration manifold $M = T^n \times I^n$, as $v^i \equiv v^i(x^i, t) := \dot{x}^i(x^i, t)$.

Note that in general, a *vector-field* represents a field of vectors defined at every point x^i within some region U (e.g., movable segments/joints only) of the total configuration manifold M (consisting of all the segments/joints). Analytically, vector-field is defined as a set of autonomous ODEs (in our case, the set (2.29)). Its solution gives the *flow*, consisting of *integral curves*

of the vector-field, such that all the vectors from the vector-field are tangent to integral curves at different representative points $x^i \in U$. In this way, through every representative point $x^i \in U$ passes both a curve from the flow and its tangent vector from the vector-field. Geometrically, vector-field is defined as a cross-section of the tangent bundle TM , the so-called *velocity phase-space*. Its *geometrical dual* is the *1-form-field*, which represents a field of one-forms (see Figure 2.1), defined at the same representative points $x^i \in U$. Analytically, 1-form-field is defined as an *exterior differential system*, an algebraic dual to the autonomous set of ODEs. Geometrically, it is defined as a cross-section of the cotangent bundle T^*M , the so-called *momentum phase-space*. Together, the vector-field and its corresponding 1-form-field define the *scalar potential field* (e.g., kinetic and/or potential energy) at the same movable region $U \subset M$.

Next, we need to formulate the internal acceleration vector-field, $a^i \equiv a^i(x^i, \dot{x}^i, t)$, acting in all movable joints, and at the same time generalizing the Newtonian 3D acceleration vector \mathbf{a} .

According to Newton, acceleration is a rate-of-change of velocity. But, from the previous subsections, we know that $a^i \neq \dot{v}^i$. However,

$$a^i := \dot{\dot{v}}^i = \dot{v}^i + \Gamma_{jk}^i v^j v^k = \ddot{x}^i + \Gamma_{jk}^i \dot{x}^j \dot{x}^k. \quad (2.30)$$

Once we have the *internal acceleration vector-field* $a^i = a^i(x^i, \dot{x}^i, t)$, defined by the set of ODEs (2.30) (including *Levi-Civita connections* Γ_{jk}^i of the Riemannian configuration manifold M), we can finally define the *internal force 1-form field*, $F_i = F_i(x^i, \dot{x}^i, t)$, as a family of force one-forms, half of them rotational and half translational, acting in all movable joints,

$$F_i := mg_{ij} a^j = mg_{ij} (\dot{v}^j + \Gamma_{ik}^j v^i v^k) = mg_{ij} (\ddot{x}^j + \Gamma_{ik}^j \dot{x}^i \dot{x}^k), \quad (2.31)$$

where we have used the simplified *material metric tensor*, mg_{ij} , for the system (considering, for simplicity, all segments to have equal mass m), defined by its *Riemannian kinetic energy form*

$$T = \frac{1}{2} mg_{ij} v^i v^j.$$

Equation $F_i = mg_{ij} a^j$, defined properly by (2.31) at every representative point x^i of the system's configuration manifold M , formulates the sought for *covariant force law*, that generalizes the fundamental Newtonian equation, $\mathbf{F} = m\mathbf{a}$, for the generic physical or engineering system. Its

meaning is:

Force 1–form–field = Mass distribution \times Acceleration vector–field

In other words, the field (or, family) of force one–forms F_i , acting in all movable joints (with constrained rotations on T^n and very restricted translations on I^n), causes both rotational and translational accelerations of all body segments, within the mass distribution mg_{ij} ⁵, along the flow–lines of the vector–field a^j .

2.1.5 Application: *Nonlinear Fluid Dynamics*

In this subsection we will derive the general form of the Navier–Stokes equations in nonlinear fluid dynamics.

2.1.5.1 Continuity Equation

Recall that the most important equation in fluid dynamics, as well as in general continuum mechanics, is the celebrated *equation of continuity*, (we explain the symbols in the following text)

$$\partial_t \rho + \operatorname{div}(\rho \mathbf{u}) = 0. \quad (2.32)$$

As a warm–up for turbulence, we will derive the continuity equation (2.32), starting from the *mass conservation principle*. Let dm denote an infinitesimal mass of a fluid particle. Then, using the absolute time derivative operator $\dot{(\quad)} \equiv \frac{D}{dt}$, the mass conservation principle reads

$$\dot{dm} = 0. \quad (2.33)$$

If we further introduce the fluid density $\rho = dm/dv$, where dv is an infinitesimal volume of a fluid particle, then the mass conservation principle (2.33) can be rewritten as

$$\dot{\rho} dv = 0,$$

⁵More realistically, instead of the simplified metric mg_{ij} we have the *material metric tensor* G_{ij} (2.21), including all k segmental masses m_χ , as well as the corresponding moments and products of inertia,

$$G_{ij}(x, m) = \sum_{\chi=1}^k m_\chi \delta_{rs} \frac{\partial y^r}{\partial x^i} \frac{\partial y^s}{\partial x^j}, \quad (r, s = 1, \dots, 6; i, j = 1, \dots, 2n).$$

which is the absolute derivative of a product, and therefore expands into

$$\dot{\rho} dv + \rho \dot{\overline{dv}} = 0. \quad (2.34)$$

Now, as the fluid density $\rho = \rho(x^k, t)$ is a function of both time t and spatial coordinates x^k , for $k = 1, 2, 3$, that is, a *scalar-field*, its total time derivative $\dot{\rho}$, figuring in (2.34), is defined by

$$\begin{aligned} \dot{\rho} &= \partial_t \rho + \partial_{x^k} \rho \partial_t x^k \equiv \partial_t \rho + \rho_{;k} u^k, \\ \text{or, in vector form} \quad \dot{\rho} &= \partial_t \rho + \text{grad}(\rho) \cdot \mathbf{u}, \end{aligned} \quad (2.35)$$

where $u^k = u^k(x^k, t) \equiv \mathbf{u}$ is the velocity vector-field of the fluid.

Regarding $\dot{\overline{dv}}$, the other term figuring in (2.34), we start by expanding an elementary volume dv along the sides $\{dx_{(p)}^i, dx_{(q)}^j, dx_{(r)}^k\}$ of an elementary parallelepiped, as

$$dv = \frac{1}{3!} \delta_{ijk}^{pqr} dx_{(p)}^i dx_{(q)}^j dx_{(r)}^k, \quad (i, j, k, p, q, r = 1, 2, 3)$$

so that its absolute derivative becomes

$$\begin{aligned} \dot{\overline{dv}} &= \frac{1}{2!} \delta_{ijk}^{pqr} \dot{\overline{dx_{(p)}^i}} dx_{(q)}^j dx_{(r)}^k \\ &= \frac{1}{2!} u_{;l}^i \delta_{ijk}^{pqr} dx_{(p)}^l dx_{(q)}^j dx_{(r)}^k \quad (\text{using } \dot{\overline{dx_{(p)}^i}} = u_{;l}^i dx_{(p)}^l), \end{aligned}$$

which finally simplifies into

$$\dot{\overline{dv}} = u_{;k}^k dv \equiv \text{div}(\mathbf{u}) dv. \quad (2.36)$$

Substituting (2.35) and (2.36) into (2.34) gives

$$\dot{\overline{\rho dv}} \equiv (\partial_t \rho + \rho_{;k} u^k) dv + \rho u_{;k}^k dv = 0. \quad (2.37)$$

As we are dealing with arbitrary fluid particles, $dv \neq 0$, so from (2.37) follows

$$\partial_t \rho + \rho_{;k} u^k + \rho u_{;k}^k \equiv \partial_t \rho + (\rho u^k)_{;k} = 0. \quad (2.38)$$

Equation (2.38) is the covariant form of the continuity equation, which in standard vector notation becomes (2.32), i.e., $\partial_t \rho + \text{div}(\rho \mathbf{u}) = 0$.

2.1.5.2 Forces Acting on a Fluid

A fluid contained in a finite volume is subject to the action of both *volume forces* F^i and *surface forces* S^i , which are respectively defined by

$$F^i = \int_v \rho f^i dv, \quad \text{and} \quad S^i = \oint_a \sigma^{ij} da_j. \quad (2.39)$$

Here, f^i is a force vector acting on an elementary mass dm , so that the elementary volume force is given by

$$dF^i = f^i dm = \rho f^i dv,$$

which is the integrand in the volume integral on l.h.s of (2.39). $\sigma^{ij} = \sigma^{ij}(x^k, t)$ is the stress tensor-field of the fluid, so that the elementary force acting on the closed oriented surface a is given by

$$dS^i = \sigma^{ij} da_j,$$

where da_j is an oriented element of the surface a ; this is the integrand in the surface integral on the r.h.s of (2.39).

On the other hand, the elementary momentum dK^i of a fluid particle (with elementary volume dv and elementary mass $dm = \rho dv$) equals the product of dm with the particle's velocity u^i , i.e.,

$$dK^i = u^i dm = \rho u^i dv,$$

so that the total momentum of the finite fluid volume v is given by the volume integral

$$K^i = \int_v \rho u^i dv. \quad (2.40)$$

Now, the *Newtonian-like force law for the fluid* states that the time derivative of the fluid momentum equals the resulting force acting on it, $\dot{K}^i = \mathcal{F}^i$, where the resulting force \mathcal{F}^i is given by the sum of surface and volume forces,

$$\mathcal{F}^i = S^i + F^i = \oint_a \sigma^{ij} da_j + \int_v \rho f^i dv. \quad (2.41)$$

From (2.40), taking the time derivative and using $\overline{\rho dv} = 0$, we get

$$\dot{K}^i = \int_v \rho \dot{u}^i dv,$$

where $\dot{u}^i = \dot{u}^i(x^k, t) \equiv \dot{\mathbf{u}}$ is the acceleration vector-field of the fluid, so that (2.41) gives

$$\oint_a \sigma^{ij} da_j + \int_v \rho(f^i - \dot{u}^i) dv = 0. \quad (2.42)$$

Now, assuming that the stress tensor $\sigma^{ij} = \sigma^{ij}(x^k, t)$ does not have any singular points in the volume v bounded by the closed surface a , we can transform the surface integral in (2.42) in the volume one, i.e.,

$$\oint_a \sigma^{ij} da_j = \int_v \sigma^{ij}_{;j} dv, \quad (2.43)$$

where $\sigma^{ij}_{;j}$ denotes the *divergence of the stress tensor*. The expression (2.43) shows us that the resulting surface force acting on the closed surface a equals the flux of the stress tensor through the surface a . Using this expression, we can rewrite (2.42) in the form

$$\int_v (\sigma^{ij}_{;j} + \rho f^i - \rho \dot{u}^i) dv = 0.$$

As this equation needs to hold for an arbitrary fluid element $dv \neq 0$, it implies the dynamical equation of motion for the fluid particles, also called the *first Cauchy law of motion*,

$$\sigma^{ij}_{;j} + \rho f^i = \rho \dot{u}^i. \quad (2.44)$$

2.1.5.3 Constitutive and Dynamical Equations

Recall that, in case of a homogenous isotropic viscous fluid, the stress tensor σ^{ij} depends on the strain-rate tensor-field $e^{ij} = e^{ij}(x^k, t)$ of the fluid in such a way that

$$\sigma^{ij} = -p g^{ij}, \quad \text{when} \quad e^{ij} = 0,$$

where the scalar function $p = p(x^k, t)$ represents the pressure field. Therefore, pressure is independent on the strain-rate tensor e^{ij} . Next, we introduce the viscosity tensor-field $\beta^{ij} = \beta^{ij}(x^k, t)$, as

$$\beta^{ij} = \sigma^{ij} + p g^{ij}, \quad (2.45)$$

which depends exclusively on the strain-rate tensor (i.e., $\beta^{ij} = 0$ whenever $e^{ij} = 0$). A viscous fluid in which the viscosity tensor β^{ij} can be expressed

as a function of the strain-rate tensor e^{ij} in the form

$$\beta^{ij} = \alpha_1(e_I, e_{II}, e_{III})g^{ij} + \alpha_2(e_I, e_{II}, e_{III})e^{ij} + \alpha_3(e_I, e_{II}, e_{III})e_k^i e^{kj}, \quad (2.46)$$

where $\alpha_l = \alpha_l(e_I, e_{II}, e_{III})$, ($l = 1, 2, 3$) are scalar functions of the basic invariants (e_I, e_{II}, e_{III}) of the strain-rate tensor e^{ij} , is called the *Stokes fluid*.

If we take only the linear terms in (2.46), we get the constitutive equation for the *Newtonian fluid*,

$$\beta^{ij} = \alpha_1 e_I g^{ij} + \alpha_2 e^{ij}, \quad (2.47)$$

which is, therefore, a linear approximation of the constitutive equation (2.46) for the Stokes fluid.

If we now put (2.47) into (2.45) we get the dynamical equation for the Newtonian fluid,

$$\sigma^{ij} = -pg^{ij} + \mu e_I g^{ij} + 2\eta e^{ij}, \quad (2.48)$$

If we put $\mu = \eta_V - \frac{2}{3}\eta$, where η_V is called the *volume viscosity* coefficient, while η is called the *shear viscosity* coefficient, we can rewrite (2.48) as

$$\sigma^{ij} = -pg^{ij} + \left(\eta_V - \frac{2}{3}\eta\right) e_I g^{ij} + 2\eta e^{ij}. \quad (2.49)$$

2.1.5.4 Navier–Stokes Equations

From the constitutive equation of the Newtonian viscous fluid (2.49), by taking the divergence, we get

$$\sigma^{ij}_{;j} = -p_{;j}g^{ij} + \left(\eta_V - \frac{2}{3}\eta\right) e_{I;j}g^{ij} + 2\eta e^{ij}_{;j}.$$

However, as $e_{I;j} = u^k_{;kj}$ as well as

$$e^{ij}_{;j} = \frac{1}{2}(u^{i;j} + u^{j;i})_{;j} = \frac{1}{2}(u^{i;j}_{;j} + u^{j;i}_{;j}) = \frac{1}{2}\Delta u^i + \frac{1}{2}u^{ki}_{;k}$$

we get

$$\sigma_{;j}^{ij} = -p_{;j}g^{ij} + \left(\eta_V - \frac{2}{3}\eta\right)u_{;kj}^kg^{ij} + \eta\Delta u^i + \eta u_{;kj}^kg^{ij},$$

or

$$\sigma_{;j}^{ij} = -p_{;j}g^{ij} + \left(\eta_V - \frac{1}{3}\eta\right)u_{;kj}^kg^{ij} + \eta\Delta u^i.$$

If we now substitute this expression into (2.44) we get

$$\rho\dot{\mathbf{u}}^i = \rho f^i - p_{;j}g^{ij} + \left(\eta_V - \frac{1}{3}\eta\right)u_{;kj}^kg^{ij} + \eta\Delta u^i, \quad (2.50)$$

that is a system of 3 scalar PDEs called the *Navier–Stokes equations*, which in vector form read

$$\rho\dot{\mathbf{u}} = \rho\mathbf{f} - \text{grad } p + \left(\eta_V - \frac{1}{3}\eta\right)\text{grad}(\text{div } \mathbf{u}) + \eta\Delta\mathbf{u}. \quad (2.51)$$

In particular, for incompressible fluids, $\text{div } \mathbf{u} = 0$, we have

$$\dot{\mathbf{u}} = \mathbf{f} - \frac{1}{\rho}\text{grad } p + \nu\Delta\mathbf{u}, \quad \text{where} \quad \nu = \frac{\eta}{\rho} \quad (2.52)$$

is the coefficient of kinematic viscosity.

2.2 Actions: The Core Machinery of Modern Physics

It is now well-known that a contemporary development of theoretical physics progresses according to the *heuristic action paradigm*⁶ (see e.g., [Ramond (1990); Feynman and Hibbs (1965); Siegel (2002)]), which follows the common three essential steps (note that many technical details are omitted here for brevity):

- (1) In order to develop a new physical theory, we first define a new *action* $\mathcal{A}[w]$, a *functional* in N *system variables* w^i , as a time integral from the *initial* point t_0 to the *final* point t_1 ,

$$\mathcal{A}[w] = \int_{t_0}^{t_1} \mathfrak{L}[w] dt. \quad (2.53)$$

⁶The action principle is a fundamental concept in physics (of as great importance as symmetry). It is very powerful for classical physics, allowing all field equations to be derived from a single function, and making symmetries simpler to check. In quantum physics the dynamics is necessarily formulated in terms of an action (in the path-integral approach), or an equivalent Hamiltonian (in the Heisenberg and Schrödinger approaches).

Now, the nature of the integrand $\mathfrak{L}[w]$ in (6.230) depends on whether we are dealing with particles or fields. In case of *particles*, $\mathfrak{L}[w] = L(q, \dot{q})$ is an ordinary finite-dimensional mechanical *Lagrangian* (usually kinetic minus potential energy), defined through mechanical (total system energy) *Hamiltonian* $H = H(q, p)$ as $\mathfrak{L}[w] = L(q, \dot{q}) = p_i \dot{q}^i - H(q, p)$, where q, \dot{q}, p are generalized coordinates, velocities and canonical momenta, respectively.

In case of *fields*, the integrand $\mathfrak{L}[w]$ is more involved, as fields just infinite-dimensional particles. Thus,

$$\mathfrak{L}[w] = \int d^n q \mathcal{L}(\varphi, \dot{\varphi}, \varphi_q),$$

where the integral is taken over all n space coordinates⁷, while $\varphi, \dot{\varphi}, \varphi_q$ denote field variables, their velocities and their coordinate (partial) derivatives, respectively. The subintegrand $\mathcal{L} = \mathcal{L}(\varphi, \dot{\varphi}, \varphi_q)$ is the system *Lagrangian density*, defined through the system *Hamiltonian density* $\mathcal{H} = \mathcal{H}(\varphi, \pi, \pi_q)$ as $\mathcal{L}(\varphi, \dot{\varphi}, \varphi_q) = \pi_i \dot{\varphi}^i - \mathcal{H}(\varphi, \pi, \pi_q)$, where π, π_q are field (canonical) momenta, and their coordinate derivatives.

- (2) Vary the action $\mathcal{A}[w]$ using the *extremal* (least) *action principle*

$$\delta \mathcal{A}[w] = 0, \quad (2.54)$$

and using techniques from *calculus of variations* (see, e.g., [Arfken (1985); Fox (1988); Ramond (1990)]), derive classical field and motion equations, as *Euler-Lagrangian equations*, describing the *extremal path*, or *direct system path*: $t \mapsto w(t)$, from t_0 to t_1 .

Again, we have two cases. The particle Euler-Lagrangian equation reads

$$\partial_t L_{\dot{q}^i} = L_{q^i},$$

and can be recast in Hamiltonian form, using the *Poisson bracket* (or, classical commutator)⁸, as a pair of canonical equations

$$\dot{q}^i = [q^i, H], \quad \dot{p}_i = [p_i, H]. \quad (2.55)$$

⁷In particular, $n = 3$ for the fields in Euclidean 3D space.

⁸Recall that for any two functions $A = A(q^k, p_k, t)$ and $B = B(q^k, p_k, t)$, their Poisson bracket is defined as

$$[A, B]_{partcl} = \left(\frac{\partial A}{\partial q^k} \frac{\partial B}{\partial p_k} - \frac{\partial A}{\partial p_k} \frac{\partial B}{\partial q^k} \right).$$

The field Euler–Lagrangian equation

$$\partial_q \mathcal{L}_{\partial_q \varphi^i} = \mathcal{L}_{\varphi^i},$$

in Hamiltonian form gives a pair of field canonical equations⁹

$$\dot{\varphi}^i = [\varphi^i, \mathcal{H}], \quad \dot{\pi}_i = [\pi_i, \mathcal{H}]. \quad (2.56)$$

- (3) Once we have a satisfactory description of fields and motions, we can perform the *Feynman quantization*¹⁰ of classical equations [Feynman and Hibbs (1965)], using the same action $\mathcal{A}[w]$ as given by (6.230), but now including *all trajectories* rather than just the extremal one. Namely, to get the *probability amplitude* $\langle f|i \rangle$ of the system *transition* from *initial state* $i(w(t_0))$ at time t_0 to *final state* $f(w(t_1))$ at time t_1 ,

⁹Here the field Poisson brackets are slightly generalized in the sense that partial derivatives ∂ are replaced with the corresponding variational derivatives δ , i.e.,

$$[A, B]_{field} = \left(\frac{\delta A}{\delta q^k} \frac{\delta B}{\delta p_k} - \frac{\delta A}{\delta p_k} \frac{\delta B}{\delta q^k} \right).$$

¹⁰Recall that quantum systems have two modes of evolution in time. The first, governed by Schrödinger equation:

$$i\hbar \frac{\partial}{\partial t} |\psi\rangle = \hat{H} |\psi\rangle,$$

(where \hat{H} is the Hamiltonian (energy) operator, $i = \sqrt{-1}$ and \hbar is Planck's constant divided by 2π ($\equiv 1$ in natural units)), describes the time evolution of quantum systems when they are undisturbed by measurements. ‘Measurements’ are defined as interactions of the system with its environment. As long as the system is sufficiently isolated from the environment, it follows Schrödinger equation. If an interaction with the environment takes place, i.e., a measurement is performed, the system abruptly decoheres i.e., collapses or reduces to one of its classically allowed states.

A *time-dependent state of a quantum system* is determined by a normalized, complex, *wave psi-function* $\psi = \psi(t)$, that is a solution of the above Schrödinger equation. In Dirac's words, this is a unit ‘ket’ vector $|\psi\rangle$ (that makes a scalar product ‘bracket’ \langle, \rangle with the dual, ‘bra’ vector $\langle\psi|$), which is an element of the *Hilbert space* $L^2(\psi)$ with a coordinate basis (q^i). The state ket–vector $|\psi(t)\rangle$ is subject to action of the *Hermitian operators* (or, self-adjoint operators), obtained by the procedure of *quantization* of classical mechanical quantities, and whose real eigen–values are being measured. *Quantum superposition* is a generalization of the algebraic principle of linear combination of vectors.

The (first) *quantization* can be performed in three different *quantum evolution pictures*, namely Schrödinger (S)–picture, in which the system state vector $|\psi(t)\rangle$ rotates and the coordinate basis (q^i) is fixed; Heisenberg (H)–picture, in which the coordinate basis rotates and the state vector is fixed; and Dirac interaction (I)–picture, in which both the state vector and the coordinate basis rotate.

we put the action $\mathcal{A}[w]$ into the *path integral*¹¹ (see, e.g., [Feynman and Hibbs (1965); Schulman (1981); Siegel (2002)]), symbolically written as

$$\langle f|i \rangle = \int_{\Omega} \mathcal{D}[w] e^{\mathbf{i}\mathcal{A}[w]}, \quad (2.57)$$

where $\mathbf{i} = \sqrt{-1}$ is the imaginary unit, Ω represents the space of *all* system paths $w^i(t)$ which contribute to the system transition with *equal probabilities*, and the implicit Planck constant \hbar is normalized to unity. This ‘functional integral’ is usually calculated by breaking up the time interval $[t_0, t_1]$ into discrete points and taking the continuum limit. The symbolic differential $\mathcal{D}[w]$ in the path integral (2.57) representing a (somewhat non-rigorous) *path measure*, defines a product

$$\mathcal{D}[w] = \prod_{i=1}^N dw^i,$$

which in case of *quantum-mechanical particles* reads

$$\mathcal{D}[w] = \prod_{i=1}^N dq^i dp_i,$$

and in case of *quantum fields* reads

$$\mathcal{D}[w] = \prod_{i=1}^N d\varphi^i d\pi_i.$$

The path integral scheme (2.57) is commonly used for calculating the *propagator* of an arbitrary quantum(gravity) system, as well as that

¹¹The *Feynman path integral* is an expression for the system’s *propagator* in terms of an integral over an infinite-dimensional *space of paths* within the system’s configuration space. It constitutes a formulation of non-relativistic quantum mechanics which is alternative to the usual Schrödinger equation. Whereas the Schrödinger equation is based on Hamiltonians, the Feynman path integral is based on Lagrangians. In the last three decades, path integrals have proven to be invaluable in quantum field theory, statistical mechanics, condensed matter physics, and quantum gravity. The path integral is now the preferred method for quantizing gauge fields, as well as setting up perturbation expansions in quantum field theory. It also leads very quickly to important conclusions in certain problems. However, for most simple non-relativistic quantum problems, the path integral is not as easy to use as the Schrödinger equation, and most of the results obtained with it can be obtained more easily by other means. Nevertheless, one cannot help but be impressed with the elegance and beauty of the Feynman path integral, or recognize that it is a result of fundamental scientific importance.

of a *Markov–Gaussian stochastic systems* described by either *Langevin* rate ODEs or *Fokker–Planck* PDEs.

The most general quantization method, the path integral (2.57), can be reduced to the most common *Dirac quantization rule* [Dirac (1982)], which uses modified particle equation (2.55) for quantum particles,

$$\dot{\hat{q}}^i = \mathbf{i}\{\hat{H}, \hat{q}^i\}, \quad \dot{\hat{p}}_i = \mathbf{i}\{\hat{H}, \hat{p}_i\}, \quad (2.58)$$

and modified field equation (6.147) for quantum fields,

$$\dot{\hat{\varphi}}^i = \mathbf{i}\{\hat{\mathcal{H}}, \hat{\varphi}^i\}, \quad \dot{\hat{\pi}}_i = \mathbf{i}\{\hat{\mathcal{H}}, \hat{\pi}_i\}, \quad (2.59)$$

where coordinate and field variables (of ordinary Euclidean space) are replaced by the corresponding *Hermitian operators* (i.e., self-adjoint operators) in the complex *Hilbert space* and the Poisson bracket $[\cdot, \cdot]$ is replaced by the quantum commutator $\{\cdot, \cdot\}$ (multiplied by $-\mathbf{i}$). In addition, the Dirac rule postulates the *Heisenberg uncertainty relations* between the canonical pairs of coordinate and field variables, namely

$$\Delta q^i \cdot \Delta p_i \geq \frac{1}{2} \quad \text{and} \quad \Delta \varphi^i \cdot \Delta \pi_i \geq \frac{1}{2}.$$

The *action paradigm*, as outlined above, provides both classical¹² and quantum description for any new physical theory, even for those yet to be discovered. It represents a heuristic tool in search for a unified force of nature; the latest theory described in this way has been the celebrated superstring theory.

2.3 Functors: Global Machinery of Modern Mathematics

In modern mathematical sciences whenever one defines a new class of mathematical objects, one proceeds almost in the next breath to say what kinds of maps between objects will be considered [Switzer (1975)]. A general framework for dealing with situations where we have some *objects* and *maps between objects*, like sets and functions, vector spaces and linear operators, points in a space and paths between points, etc. – gives the modern meta-language of categories and functors. Categories are mathematical universes and functors are ‘projectors’ from one universe onto another. For this rea-

¹²For example, the *Einstein–Hilbert action* is used to derive the *Einstein equation* of general relativity (see below).

son, in this book we extensively use this language, mainly following its founder, S. MacLane [MacLane (1971)].

2.3.1 Maps

2.3.1.1 Notes from Set Theory

Given a map (or, a function) $f : A \rightarrow B$, the set A is called the *domain* of f , and denoted $\text{Dom } f$. The set B is called the *codomain* of f , and denoted $\text{Cod } f$. The codomain is not to be confused with the *range* of $f(A)$, which is in general only a subset of B .

A map $f : X \rightarrow Y$ is called *injective* or 1–1 or an *injection* if for every y in the codomain Y there is at most one x in the domain X with $f(x) = y$. Put another way, given x and x' in X , if $f(x) = f(x')$, then it follows that $x = x'$. A map $f : X \rightarrow Y$ is called *surjective* or *onto* or a *surjection* if for every y in the codomain $\text{Cod } f$ there is at least one x in the domain X with $f(x) = y$. Put another way, the *range* $f(X)$ is equal to the codomain Y . A map is *bijective* iff it is both injective and surjective. Injective functions are called the *monomorphisms*, and surjective functions are called the *epimorphisms* in the *category of sets* (see below).

Two main classes of maps (or, functions) that we will use in this book are: (i) continuous maps (denoted as C^0 –class), and (ii) smooth or differentiable maps (denoted as C^k –class). The former class is the core of topology, the latter of differential geometry. They are both used in the core concept of manifold.

A *relation* is any subset of a *Cartesian product* (see below). By definition, an *equivalence relation* α on a set X is a relation which is *reflexive*, *symmetrical* and *transitive*, i.e., relation that satisfies the following three conditions:

- (1) *Reflexivity*: each element $x \in X$ is equivalent to itself, i.e., $x\alpha x$,
- (2) *Symmetry*: for any two elements $x, x' \in X$, $x\alpha x'$ implies $x'\alpha x$, and
- (3) *Transitivity*: $a \leq b$ and $b \leq c$ implies $a \leq c$.

Similarly, a relation \leq defines a *partial order* on a set S if it has the following properties:

- (1) *Reflexivity*: $a \leq a$ for all $a \in S$,
- (2) *Antisymmetry*: $a \leq b$ and $b \leq a$ implies $a = b$, and
- (3) *Transitivity*: $a \leq b$ and $b \leq c$ implies $a \leq c$.

A *partially ordered set* (or *poset*) is a set taken together with a partial order on it. Formally, a partially ordered set is defined as an ordered pair $P = (X, \leq)$, where X is called the *ground set* of P and \leq is the partial order of P .

2.3.1.2 Notes From Calculus

2.3.1.3 Maps

Recall that a *map* (or, *function*) f is a *rule* that assigns to each element x in a set A exactly one element, called $f(x)$, in a set B . A map could be thought of as a *machine* $[[f]]$ with x -input (the *domain* of f is the set of all possible inputs) and $f(x)$ -output (the *range* of f is the set of all possible outputs) [Stuart (1999)]

$$x \rightarrow [[f]] \rightarrow f(x)$$

There are four possible ways to represent a function (or map): (i) verbally (by a description in words); (ii) numerically (by a table of values); (iii) visually (by a graph); and (iv) algebraically (by an explicit formula). The most common method for visualizing a function is its *graph*. If f is a function with domain A , then its graph is the set of ordered input-output pairs

$$\{(x, f(x)) : x \in A\}.$$

A generalization of the graph concept is a concept of a *cross-section of a fibre bundle*, which is one of the core geometrical objects for dynamics of complex systems.

2.3.1.4 Algebra of Maps

Let f and g be maps with domains A and B . Then the maps $f + g$, $f - g$, fg , and f/g are defined as follows [Stuart (1999)]

$$\begin{aligned} (f + g)(x) &= f(x) + g(x) & \text{domain} &= A \cap B, \\ (f - g)(x) &= f(x) - g(x) & \text{domain} &= A \cap B, \\ (fg)(x) &= f(x)g(x) & \text{domain} &= A \cap B, \\ \left(\frac{f}{g}\right)(x) &= \frac{f(x)}{g(x)} & \text{domain} &= \{x \in A \cap B : g(x) \neq 0\}. \end{aligned}$$

2.3.1.5 Compositions of Maps

Given two maps f and g , the composite map $f \circ g$ (also called the *composition* of f and g) is defined by

$$(f \circ g)(x) = f(g(x)).$$

The $(f \circ g)$ -machine is composed of the g -machine (first) and then the f -machine [Stuart (1999)],

$$x \rightarrow [[g]] \rightarrow g(x) \rightarrow [[f]] \rightarrow f(g(x))$$

For example, suppose that $y = f(u) = \sqrt{u}$ and $u = g(x) = x^2 + 1$. Since y is a function of u and u is a function of x , it follows that y is ultimately a function of x . We calculate this by substitution

$$y = f(u) = f \circ g = f(g(x)) = f(x^2 + 1) = \sqrt{x^2 + 1}.$$

2.3.1.6 The Chain Rule

If f and g are both differentiable (or smooth, i.e., C^k) maps and $h = f \circ g$ is the composite map defined by $h(x) = f(g(x))$, then h is differentiable and h' is given by the product [Stuart (1999)]

$$h'(x) = f'(g(x)) g'(x).$$

In Leibniz notation, if $y = f(u)$ and $u = g(x)$ are both differentiable maps, then

$$\frac{dy}{dx} = \frac{dy}{du} \frac{du}{dx}.$$

The reason for the name *chain rule* becomes clear if we add another link to the chain. Suppose that we have one more differentiable map $x = h(t)$. Then, to calculate the derivative of y with respect to t , we use the chain rule twice,

$$\frac{dy}{dt} = \frac{dy}{du} \frac{du}{dx} \frac{dx}{dt}.$$

2.3.1.7 Integration and Change of Variables

1–1 continuous (i.e., C^0) map T with a nonzero *Jacobian* $\left| \frac{\partial(x, \dots)}{\partial(u, \dots)} \right|$ that maps a region S onto a region R , (see [Stuart (1999)]) we have the following substitution formulas:

1. for a single integral,

$$\int_R f(x) dx = \int_S f(x(u)) \frac{\partial x}{\partial u} du,$$

2. for a double integral,

$$\iint_R f(x, y) dA = \iint_S f(x(u, v), y(u, v)) \left| \frac{\partial(x, y)}{\partial(u, v)} \right| dudv,$$

3. for a triple integral,

$$\iiint_R f(x, y, z) dV = \iiint_S f(x(u, v, w), y(u, v, w), z(u, v, w)) \left| \frac{\partial(x, y, z)}{\partial(u, v, w)} \right| dudvdw$$

4. similarly for n -tuple integrals.

2.3.1.8 Notes from General Topology

Topology is a kind of *abstraction* of Euclidean geometry, and also a natural framework for the study of *continuity*.¹³ Euclidean geometry is abstracted by regarding triangles, circles, and squares as being the same basic object. Continuity enters because in saying this one has in mind a *continuous deformation* of a triangle into a square or a circle, or any arbitrary shape. On the other hand, a disk with a hole in the center is topologically different from a circle or a square because one cannot create or destroy holes by continuous deformations. Thus using topological methods one does not expect to be able to identify a geometrical figure as being a triangle or a square. However, one does expect to be able to detect the presence of gross features such as holes or the fact that the figure is made up of two disjoint pieces etc. In this way topology produces theorems that are usually qualitative in nature – they may assert, for example, the existence or non-existence of an object. They will not, in general, give the means for its construction [Nash and Sen (1983)].

¹³Intuitively speaking, a function $f : \mathbb{R} \rightarrow \mathbb{R}$ is continuous near a point x in its domain if its value does not jump there. That is, if we just take δx to be small enough, the two function values $f(x)$ and $f(x + \delta x)$ should approach each other arbitrarily closely. In more rigorous terms, this leads to the following definition: A function $f : \mathbb{R} \rightarrow \mathbb{R}$ is continuous at $x \in \mathbb{R}$ if for all $\epsilon > 0$, there exists a $\delta > 0$ such that for all $y \in \mathbb{R}$ with $|y - x| < \delta$, we have that $|f(y) - f(x)| < \epsilon$. The whole function is called continuous if it is continuous at every point x .

2.3.1.9 Topological Space

Study of topology starts with the fundamental notion of *topological space*. Let X be any *set* and $Y = \{X_\alpha\}$ denote a collection, finite or infinite of subsets of X . Then X and Y form a topological space provided the X_α and Y satisfy:

- (1) Any finite or infinite subcollection $\{Z_\alpha\} \subset X_\alpha$ has the property that $\cup Z_\alpha \in Y$, and
- (2) Any *finite subcollection* $\{Z_{\alpha_1}, \dots, Z_{\alpha_n}\} \subset X_\alpha$ has the property that $\cap Z_{\alpha_i} \in Y$.

The set X is then called a topological space and the X_α are called *open sets*. The choice of Y satisfying (2) is said to give a topology to X .

Given two topological spaces X and Y , a *function* (or, a *map*) $f : X \rightarrow Y$ is *continuous* if the inverse image of an open set in Y is an open set in X .

The main general idea in topology is to study spaces which can be continuously deformed into one another, namely the idea of *homeomorphism*. If we have two topological spaces X and Y , then a map $f : X \rightarrow Y$ is called a homeomorphism iff

- (1) f is continuous (C^0), and
- (2) There exists an inverse of f , denoted f^{-1} , which is also continuous.

Definition (2) implies that if f is a homeomorphism then so is f^{-1} . Homeomorphism is the main topological example of *reflexive*, *symmetrical* and *transitive relation*, i.e., *equivalence relation*. Homeomorphism divides all topological spaces up into *equivalence classes*. In other words, a pair of topological spaces, X and Y , belong to the same equivalence class if they are homeomorphic.

The second example of topological equivalence relation is *homotopy*. While homeomorphism generates equivalence classes whose members are topological spaces, homotopy generates equivalence classes whose members are continuous (C^0) maps. Consider two continuous maps $f, g : X \rightarrow Y$ between topological spaces X and Y . Then the map f is said to be *homotopic* to the map g if f can be continuously deformed into g (see below for the precise definition of homotopy). Homotopy is an equivalence relation which divides the space of continuous maps between two topological spaces into equivalence classes [Nash and Sen (1983)].

Another important notions in topology are *covering*, *compactness* and *connectedness*. Given a family of sets $\{X_\alpha\} = X$ say, then X is a *covering* of another set Y if $\cup X_\alpha$ contains Y . If all the X_α happen to be open sets the covering is called an *open covering*. Now consider the set Y and all its possible open coverings. The set Y is *compact* if for every open covering $\{X_\alpha\}$ with $\cup X_\alpha \supset Y$ there always exists a finite subcovering $\{X_1, \dots, X_n\}$ of Y with $X_1 \cup \dots \cup X_n \supset Y$. Again, we define a set Z to be *connected* if it cannot be written as $Z = Z_1 \cup Z_2$, where Z_1 and Z_2 are both open and $Z_1 \cap Z_2$ is an empty set.

Let A_1, A_2, \dots, A_n be closed subspaces of a topological space X such that $X = \cup_{i=1}^n A_i$. Suppose $f_i : A_i \rightarrow Y$ is a function, $1 \leq i \leq n$, iff

$$f_i|_{A_i \cap A_j} = f_j|_{A_i \cap A_j}, 1 \leq i, j \leq n. \quad (2.60)$$

In this case f is continuous iff each f_i is. Using this procedure we can define a C^0 -function $f : X \rightarrow Y$ by cutting up the space X into closed subsets A_i and defining f on each A_i separately in such a way that $f|_{A_i}$ is obviously continuous; we then have only to check that the different definitions agree on the *overlaps* $A_i \cap A_j$.

The *universal property of the Cartesian product*: let $p_X : X \times Y \rightarrow X$, and $p_Y : X \times Y \rightarrow Y$ be the *projections* onto the first and second factors, respectively. Given any pair of functions $f : Z \rightarrow X$ and $g : Z \rightarrow Y$ there is a unique function $h : Z \rightarrow X \times Y$ such that $p_X \circ h = f$, and $p_Y \circ h = g$. Function h is continuous iff both f and g are. This property characterizes X/α up to homeomorphism. In particular, to check that a given function $h : Z \rightarrow X$ is continuous it will suffice to check that $p_X \circ h$ and $p_Y \circ h$ are continuous.

The *universal property of the quotient*: let α be an equivalence relation on a topological space X , let X/α denote the *space of equivalence classes* and $p_\alpha : X \rightarrow X/\alpha$ the *natural projection*. Given a function $f : X \rightarrow Y$, there is a function $f' : X/\alpha \rightarrow Y$ with $f' \circ p_\alpha = f$ iff $x\alpha x'$ implies $f(x) = f(x')$, for all $x \in X$. In this case f' is continuous iff f is. This property characterizes X/α up to homeomorphism.

2.3.1.10 Homotopy

Now we return to the fundamental notion of homotopy. Let I be a compact unit interval $I = [0, 1]$. A *homotopy* from X to Y is a continuous function $F : X \times I \rightarrow Y$. For each $t \in I$ one has $F_t : X \rightarrow Y$ defined by $F_t(x) = F(x, t)$ for all $x \in X$. The functions F_t are called the ‘stages’ of the

homotopy. If $f, g : X \rightarrow Y$ are two continuous maps, we say f is homotopic to g , and write $f \simeq g$, if there is a homotopy $F : X \times I \rightarrow Y$ such that $F_0 = f$ and $F_1 = g$. In other words, f can be continuously deformed into g through the stages F_t . If $A \subset X$ is a subspace, then F is a homotopy relative to A if $F(a, t) = F(a, 0)$, for all $a \in A, t \in I$.

The homotopy relation \simeq is an equivalence relation. To prove that we have $f \simeq f$ is obvious; take $F(x, t) = f(x)$, for all $x \in X, t \in I$. If $f \simeq g$ and F is a homotopy from f to g , then $G : X \times I \rightarrow Y$ defined by $G(x, t) = F(x, 1 - t)$, is a homotopy from g to f , i.e., $g \simeq f$. If $f \simeq g$ with homotopy F and $g \simeq h$ with homotopy G , then $f \simeq h$ with homotopy H defined by

$$H(x, t) = \begin{cases} F(x, t), & 0 \leq t \leq 1/2 \\ G(x, 2t - 1), & 1/2 \leq t \leq 1 \end{cases}.$$

To show that H is continuous we use the relation (2.60).

In this way, the set of all C^0 -functions $f : X \rightarrow Y$ between two topological spaces X and Y , called the *function space* and denoted by Y^X , is partitioned into equivalence classes under the relation \simeq . The equivalence classes are called *homotopy classes*, the homotopy class of f is denoted by $[f]$, and the set of all homotopy classes is denoted by $[X; Y]$.

If α is an equivalence relation on a topological space X and $F : X \times I \rightarrow Y$ is a homotopy such that each stage F_t factors through X/α , i.e., $x\alpha x'$ implies $F_t(x) = F_t(x')$, then F induces a homotopy $F' : (X/\alpha) \times I \rightarrow Y$ such that $F' \circ (p_\alpha \times 1) = F$.

Homotopy theory has a range of applications of its own, outside topology and geometry, as for example in proving Cauchy Theorem in complex variable theory, or in solving nonlinear equations of artificial neural networks.

A *pointed set* (S, s_0) is a set S together with a distinguished point $s_0 \in S$. Similarly, a *pointed topological space* (X, x_0) is a space X together with a distinguished point $x_0 \in X$. When we are concerned with pointed spaces $(X, x_0), (Y, y_0)$, etc., we always require that all functions $f : X \rightarrow Y$ shall preserve base points, i.e., $f(x_0) = y_0$, and that all homotopies $F : X \times I \rightarrow Y$ be relative to the base point, i.e., $F(x_0, t) = y_0$, for all $t \in I$. We denote the homotopy classes of base point-preserving functions by $[X, x_0; Y, y_0]$ (where homotopies are relative to x_0). $[X, x_0; Y, y_0]$ is a pointed set with base point f_0 , the constant function: $f_0(x) = y_0$, for all $x \in X$.

A *path* $\gamma(t)$ from x_0 to x_1 in a topological space X is a continuous map

$\gamma : I \rightarrow X$ with $\gamma(0) = x_0$ and $\gamma(1) = x_1$. Thus X^I is the space of all paths in X with the compact-open topology. We introduce a relation \sim on X by saying $x_0 \sim x_1$ iff there is a path $\gamma : I \rightarrow X$ from x_0 to x_1 . \sim is clearly an equivalence relation, and the set of equivalence classes is denoted by $\pi_0(X)$. The elements of $\pi_0(X)$ are called the *path components*, or *0-components* of X . If $\pi_0(X)$ contains just one element, then X is called *path connected*, or *0-connected*. A *closed path*, or *loop* in X at the point x_0 is a path $\gamma(t)$ for which $\gamma(0) = \gamma(1) = x_0$. The *inverse loop* $\gamma^{-1}(t)$ based at $x_0 \in X$ is defined by $\gamma^{-1}(t) = \gamma(1 - t)$, for $0 \leq t \leq 1$. The *homotopy of loops* is the particular case of the above defined homotopy of continuous maps.

If (X, x_0) is a pointed space, then we may regard $\pi_0(X)$ as a pointed set with the 0-component of x_0 as a base point. We use the notation $\pi_0(X, x_0)$ to denote $p_0(X, x_0)$ thought of as a pointed set. If $f : X \rightarrow Y$ is a map then f sends 0-components of X into 0-components of Y and hence defines a function $\pi_0(f) : \pi_0(X) \rightarrow \pi_0(Y)$. Similarly, a base point-preserving map $f : (X, x_0) \rightarrow (Y, y_0)$ induces a map of pointed sets $\pi_0(f) : \pi_0(X, x_0) \rightarrow \pi_0(Y, y_0)$. In this way defined π_0 represents a ‘functor’ from the ‘category’ of topological (point) spaces to the underlying category of (point) sets (see the next section).

Combination of topology and calculus gives differential topology, or differential geometry.

2.3.1.11 Commutative Diagrams

The *category theory* (see below) was born with an observation that many properties of mathematical systems can be unified and simplified by a presentation with *commutative diagrams of arrows* [MacLane (1971)]. Each arrow $f : X \rightarrow Y$ represents a function (i.e., a map, transformation, operator); that is, a source (domain) set X , a target (codomain) set Y , and a rule $x \mapsto f(x)$ which assigns to each element $x \in X$ an element $f(x) \in Y$. A typical diagram of sets and functions is

$$\begin{array}{ccc}
 X & \xrightarrow{f} & Y \\
 & \searrow h & \downarrow g \\
 & & Z
 \end{array}
 \quad \text{or} \quad
 \begin{array}{ccc}
 X & \xrightarrow{f} & f(X) \\
 & \searrow h & \downarrow g \\
 & & g(f(X))
 \end{array}$$

This diagram is *commutative* iff $h = g \circ f$, where $g \circ f$ is the usual composite function $g \circ f : X \rightarrow Z$, defined by $x \mapsto g(f(x))$.

Similar commutative diagrams apply in other mathematical, physical and computing contexts; e.g., in the ‘category’ of all topological spaces, the letters X, Y , and Z represent topological spaces while f, g , and h stand for continuous maps. Again, in the category of all groups, X, Y , and Z stand for groups, f, g , and h for homomorphisms.

Less formally, composing maps is like following directed paths from one object to another (e.g., from set to set). In general, a diagram is commutative iff any two paths along arrows that start at the same point and finish at the same point yield the same ‘homomorphism’ via compositions along successive arrows. Commutativity of the whole diagram follows from commutativity of its triangular components (depicting a ‘commutative flow’, see Figure 2.4). Study of commutative diagrams is popularly called ‘diagram chasing’, and provides a powerful tool for mathematical thought.

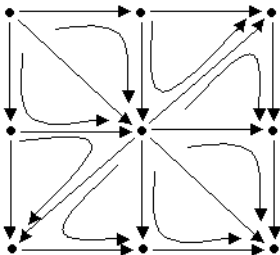
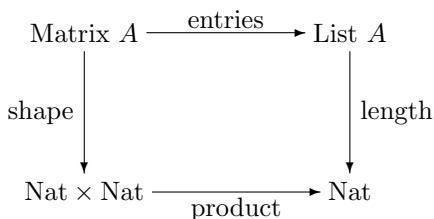


Fig. 2.4 A *commutative flow* (denoted by curved arrows) on a *triangulated digraph*. Commutativity of the whole diagram follows from commutativity of its triangular components.

As an example from linear algebra, consider an elementary diagrammatic description of matrices, using the following *pull-back diagram* [Barry (1993)]:



asserts that a matrix is determined by its shape, given by a pair of natural numbers representing the number of rows and columns, and its data, given by the matrix entries listed in some specified order.

Many properties of mathematical constructions may be represented by *universal properties* of diagrams [MacLane (1971)]. Consider the *Cartesian product* $X \times Y$ of two sets, consisting as usual of all ordered pairs $\langle x, y \rangle$ of elements $x \in X$ and $y \in Y$. The projections $\langle x, y \rangle \mapsto x$, $\langle x, y \rangle \mapsto y$ of the product on its ‘axes’ X and Y are functions $p : X \times Y \rightarrow X$, $q : X \times Y \rightarrow Y$. Any function $h : W \rightarrow X \times Y$ from a third set W is uniquely determined by its composites $p \circ h$ and $q \circ h$. Conversely, given W and two functions f and g as in the diagram below, there is a unique function h which makes the following diagram commute:

$$\begin{array}{ccccc} & & W & & \\ & \swarrow f & \downarrow h & \searrow g & \\ X & \xleftarrow{p} & X \times Y & \xrightarrow{q} & Y \end{array}$$

This property describes the Cartesian product $X \times Y$ uniquely; the same diagram, read in the category of topological spaces or of groups, describes uniquely the Cartesian product of spaces or of the direct product of groups.

The construction ‘Cartesian product’ is technically called a ‘functor’ because it applies suitably both to the sets and to the functions between them; two functions $k : X \rightarrow X'$ and $l : Y \rightarrow Y'$ have a function $k \times l$ as their Cartesian product:

$$k \times l : X \times Y \rightarrow X' \times Y', \quad \langle x, y \rangle \mapsto \langle kx, ly \rangle.$$

2.3.1.12 Groups and Related Algebraic Structures

As already stated, the basic functional unit of lower biodynamics is the special Euclidean group $SE(3)$ of rigid body motions. In general, a *group* is a pointed set (G, e) with a *multiplication* $\mu : G \times G \rightarrow G$ and an *inverse* $\nu : G \rightarrow G$ such that the following diagrams commute [Switzer (1975)]:

(1)

$$\begin{array}{ccccc} G & \xrightarrow{(e, 1)} & G \times G & \xrightarrow{(1, e)} & G \\ & \searrow 1 & \downarrow \mu & \swarrow 1 & \\ & & G & & \end{array}$$

(e is a two-sided identity)

(2)

$$\begin{array}{ccc}
 G \times G \times G & \xrightarrow{\mu \times 1} & G \times G \\
 1 \times \mu \downarrow & & \downarrow \mu \\
 G \times G & \xrightarrow{\mu} & G
 \end{array}$$

(associativity)

(3)

$$\begin{array}{ccccc}
 G & \xrightarrow{(\nu, 1)} & G \times G & \xrightarrow{(1, \nu)} & G \\
 & \searrow e & \downarrow \mu & & \nearrow e \\
 & & G & &
 \end{array}$$

(inverse).

Here $e : G \rightarrow G$ is the constant map $e(g) = e$ for all $g \in G$. $(e, 1)$ means the map such that $(e, 1)(g) = (e, g)$, etc. A group G is called *commutative* or *Abelian group* if in addition the following diagram commutes

$$\begin{array}{ccc}
 G \times G & \xrightarrow{T} & G \times G \\
 & \searrow \mu & \nearrow \mu \\
 & G &
 \end{array}$$

where $T : G \times G \rightarrow G \times G$ is the switch map $T(g_1, g_2) = (g_2, g_1)$, for all $(g_1, g_2) \in G \times G$.

A group G *acts* (on the left) on a set A if there is a function $\alpha : G \times A \rightarrow A$ such that the following diagrams commute [Switzer (1975)]:

(1)

$$\begin{array}{ccc}
 A & \xrightarrow{(e, 1)} & G \times A \\
 & \searrow 1 & \downarrow \alpha \\
 & & A
 \end{array}$$

(2)

$$\begin{array}{ccc}
 G \times G \times A & \xrightarrow{1 \times \alpha} & G \times A \\
 \mu \times 1 \downarrow & & \downarrow \alpha \\
 G \times A & \xrightarrow{\alpha} & A
 \end{array}$$

where $(e, 1)(x) = (e, x)$ for all $x \in A$. The *orbits* of the action are the sets $Gx = \{gx : g \in G\}$ for all $x \in A$.

Given two groups $(G, *)$ and (H, \cdot) , a *group homomorphism* from $(G, *)$ to (H, \cdot) is a function $h : G \rightarrow H$ such that for all x and y in G it holds that

$$h(x * y) = h(x) \cdot h(y).$$

From this property, one can deduce that h maps the identity element e_G of G to the identity element e_H of H , and it also maps inverses to inverses in the sense that $h(x^{-1}) = h(x)^{-1}$. Hence one can say that h is *compatible* with the *group structure*.

The *kernel* $\text{Ker } h$ of a group homomorphism $h : G \rightarrow H$ consists of all those elements of G which are sent by h to the identity element e_H of H , i.e.,

$$\text{Ker } h = \{x \in G : h(x) = e_H\}.$$

The *image* $\text{Im } h$ of a group homomorphism $h : G \rightarrow H$ consists of all elements of H which are sent by h to H , i.e.,

$$\text{Im } h = \{h(x) : x \in G\}.$$

The kernel is a *normal subgroup* of G and the image is a *subgroup* of H . The homomorphism h is *injective* (and called a *group monomorphism*) iff $\text{Ker } h = e_G$, i.e., iff the kernel of h consists of the identity element of G only.

Similarly, a *ring* is a set S together with two binary operators $+$ and $*$ (commonly interpreted as addition and multiplication, respectively) satisfying the following conditions:

- (1) Additive associativity: For all $a, b, c \in S$, $(a + b) + c = a + (b + c)$,
- (2) Additive commutativity: For all $a, b \in S$, $a + b = b + a$,

- (3) Additive identity: There exists an element $0 \in S$ such that for all $a \in S$, $0 + a = a + 0 = a$,
- (4) Additive inverse: For every $a \in S$ there exists $-a \in S$ such that $a + (-a) = (-a) + a = 0$,
- (5) Multiplicative associativity: For all $a, b, c \in S$, $(a * b) * c = a * (b * c)$,
- (6) Left and right distributivity: For all $a, b, c \in S$, $a * (b + c) = (a * b) + (a * c)$ and $(b + c) * a = (b * a) + (c * a)$.

A ring (the term introduced by *David Hilbert*) is therefore an Abelian group under addition and a semigroup under multiplication. A ring that is commutative under multiplication, has a unit element, and has no divisors of zero is called an *integral domain*. A ring which is also a commutative multiplication group is called a *field*. The simplest rings are the integers \mathbb{Z} , polynomials $R[x]$ and $R[x, y]$ in one and two variables, and square $n \times n$ real matrices.

An *ideal* is a subset \mathfrak{I} of elements in a ring R which forms an additive group and has the property that, whenever x belongs to R and y belongs to \mathfrak{I} , then xy and yx belong to \mathfrak{I} . For example, the set of even integers is an ideal in the ring of integers \mathbb{Z} . Given an ideal \mathfrak{I} , it is possible to define a factor ring R/\mathfrak{I} .

A ring is called *left* (respectively, *right*) *Noetherian* if it does not contain an infinite ascending chain of left (respectively, right) ideals. In this case, the ring in question is said to satisfy the ascending chain condition on left (respectively, right) ideals. A *ring* is said to be *Noetherian* if it is both left and right Noetherian. If a ring R is Noetherian, then the following are equivalent:

- (1) R satisfies the ascending chain condition on ideals.
- (2) Every ideal of R is finitely generated.
- (3) Every set of ideals contains a maximal element.

A *module* is a mathematical object in which things can be added together commutatively by multiplying coefficients and in which most of the rules of manipulating vectors hold. A module is abstractly very similar to a vector space, although in modules, coefficients are taken in rings which are much more general algebraic objects than the fields used in vector spaces. A module taking its coefficients in a ring R is called a module over R or R -module. Modules are the basic tool of homological algebra.

Examples of modules include the set of integers \mathbb{Z} , the cubic lattice in d dimensions \mathbb{Z}^d , and the group ring of a group. \mathbb{Z} is a module over

itself. It is closed under addition and subtraction. Numbers of the form $n\alpha$ for $n \in \mathbb{Z}$ and α a fixed integer form a submodule since, for $(n, m) \in \mathbb{Z}$, $n\alpha \pm m\alpha = (n \pm m)\alpha$ and $(n \pm m)$ is still in \mathbb{Z} . Also, given two integers a and b , the smallest module containing a and b is the module for their greatest common divisor, $\alpha = GCD(a, b)$.

A module M is a *Noetherian module* if it obeys the ascending chain condition with respect to inclusion, i.e., if every set of increasing sequences of submodules eventually becomes constant. If a module M is Noetherian, then the following are equivalent:

- (1) M satisfies the ascending chain condition on submodules.
- (2) Every submodule of M is finitely generated.
- (3) Every set of submodules of M contains a maximal element.

Let I be a partially ordered set. A *direct system* of R -modules over I is an ordered pair $\{M_i, \varphi_j^i\}$ consisting of an indexed family of modules $\{M_i : i \in I\}$ together with a family of homomorphisms $\{\varphi_j^i : M_i \rightarrow M_j\}$ for $i \leq j$, such that $\varphi_i^i = 1_{M_i}$ for all i and such that the following diagram commutes whenever $i \leq j \leq k$

$$\begin{array}{ccc}
 M_i & \xrightarrow{\varphi_k^i} & M_k \\
 & \searrow \varphi_k^j & \nearrow \varphi_j^i \\
 & M_j &
 \end{array}$$

Similarly, an *inverse system* of R -modules over I is an ordered pair $\{M_i, \psi_j^i\}$ consisting of an indexed family of modules $\{M_i : i \in I\}$ together with a family of homomorphisms $\{\psi_j^i : M_j \rightarrow M_i\}$ for $i \leq j$, such that $\psi_i^i = 1_{M_i}$ for all i and such that the following diagram commutes whenever $i \leq j \leq k$

$$\begin{array}{ccc}
 M_k & \xrightarrow{\psi_i^k} & M_i \\
 & \searrow \psi_j^k & \nearrow \psi_i^j \\
 & M_j &
 \end{array}$$

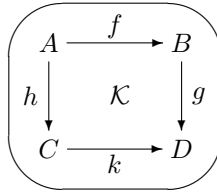
2.3.2 Categories

A category is a generic mathematical structure consisting of a collection of *objects* (sets with possibly additional structure), with a corresponding collection of *arrows*, or *morphisms*, between objects (agreeing with this additional structure). A category \mathcal{K} is defined as a pair $(\mathbf{Ob}(\mathcal{K}), \mathbf{Mor}(\mathcal{K}))$ of generic objects A, B, \dots in $\mathbf{Ob}(\mathcal{K})$ and generic arrows $f : A \rightarrow B$, $g : B \rightarrow C, \dots$ in $\mathbf{Mor}(\mathcal{K})$ between objects, with *associative composition*:

$$A \xrightarrow{f} B \xrightarrow{g} C = A \xrightarrow{f \circ g} C,$$

and *identity* (loop) arrow. (Note that in topological literature, $\mathbf{Hom}(\mathcal{K})$ or $\mathbf{hom}(\mathcal{K})$ is used instead of $\mathbf{Mor}(\mathcal{K})$; see [Switzer (1975)]).

A category \mathcal{K} is usually depicted as a *commutative diagram* (i.e., a diagram with a common *initial object* A and *final object* D):



To make this more precise, we say that a *category* \mathcal{K} is defined if we have:

- (1) A *class of objects* $\{A, B, C, \dots\}$ of \mathcal{K} , denoted by $\mathbf{Ob}(\mathcal{K})$;
- (2) A *set of morphisms*, or *arrows* $\mathbf{Mor}_{\mathcal{K}}(A, B)$, with elements $f : A \rightarrow B$, defined for any *ordered pair* $(A, B) \in \mathcal{K}$, such that for two different pairs $(A, B) \neq (C, D)$ in \mathcal{K} , we have $\mathbf{Mor}_{\mathcal{K}}(A, B) \cap \mathbf{Mor}_{\mathcal{K}}(C, D) = \emptyset$;
- (3) For any *triplet* $(A, B, C) \in \mathcal{K}$ with $f : A \rightarrow B$ and $g : B \rightarrow C$, there is a *composition* of morphisms

$$\mathbf{Mor}_{\mathcal{K}}(B, C) \times \mathbf{Mor}_{\mathcal{K}}(A, B) \ni (g, f) \rightarrow g \circ f \in \mathbf{Mor}_{\mathcal{K}}(A, C),$$

written schematically as

$$\frac{f : A \rightarrow B, \quad g : B \rightarrow C}{g \circ f : A \rightarrow C}.$$

If we have a morphism $f \in \mathbf{Mor}_{\mathcal{K}}(A, B)$, (otherwise written $f : A \rightarrow B$, or $A \xrightarrow{f} B$), then $A = \text{dom}(f)$ is a *domain* of f , and $B = \text{cod}(f)$ is a *codomain* of f (of which *range* of f is a subset) and denoted $B = \text{ran}(f)$.

To make \mathcal{K} a category, it must also fulfill the following two properties:

- (1) *Associativity of morphisms*: for all $f \in \text{Mor}_{\mathcal{K}}(A, B)$, $g \in \text{Mor}_{\mathcal{K}}(B, C)$, and $h \in \text{Mor}_{\mathcal{K}}(C, D)$, we have $h \circ (g \circ f) = (h \circ g) \circ f$; in other words, the following diagram is commutative

$$\begin{array}{ccc}
 A & \xrightarrow{h \circ (g \circ f) = (h \circ g) \circ f} & D \\
 f \downarrow & & \uparrow h \\
 B & \xrightarrow{g} & C
 \end{array}$$

- (2) *Existence of identity morphism*: for every object $A \in \text{Ob}(\mathcal{K})$ exists a unique identity morphism $1_A \in \text{Mor}_{\mathcal{K}}(A, A)$; for any two morphisms $f \in \text{Mor}_{\mathcal{K}}(A, B)$, and $g \in \text{Mor}_{\mathcal{K}}(B, C)$, compositions with identity morphism $1_B \in \text{Mor}_{\mathcal{K}}(B, B)$ give $1_B \circ f = f$ and $g \circ 1_B = g$, i.e., the following diagram is commutative:

$$\begin{array}{ccccc}
 A & \xrightarrow{f} & B & \xrightarrow{g} & C \\
 & \searrow f & \downarrow 1_B & \nearrow g & \\
 & & B & &
 \end{array}$$

The set of all morphisms of the category \mathcal{K} is denoted

$$\text{Mor}(\mathcal{K}) = \bigcup_{A, B \in \text{Ob}(\mathcal{K})} \text{Mor}_{\mathcal{K}}(A, B).$$

If for two morphisms $f \in \text{Mor}_{\mathcal{K}}(A, B)$ and $g \in \text{Mor}_{\mathcal{K}}(B, A)$ the equality $g \circ f = 1_A$ is valid, then the morphism g is said to be *left inverse* (or *retraction*), of f , and f *right inverse* (or *section*) of g . A morphism which is both right and left inverse of f is said to be *two-sided inverse* of f .

A morphism $m : A \rightarrow B$ is called *monomorphism* in \mathcal{K} (i.e., *1-1*, or *injection* map), if for any two parallel morphisms $f_1, f_2 : C \rightarrow A$ in \mathcal{K} the equality $m \circ f_1 = m \circ f_2$ implies $f_1 = f_2$; in other words, m is monomorphism if it is *left cancellable*. Any morphism with a left inverse is monomorphism.

A morphism $e : A \rightarrow B$ is called *epimorphism* in \mathcal{K} (i.e., *onto*, or *surjection* map), if for any two morphisms $g_1, g_2 : B \rightarrow C$ in \mathcal{K} the equality $g_1 \circ e = g_2 \circ e$ implies $g_1 = g_2$; in other words, e is epimorphism if it is *right cancellable*. Any morphism with a right inverse is epimorphism.

A morphism $f : A \rightarrow B$ is called *isomorphism* in \mathcal{K} (denoted as $f : A \cong B$) if there exists a morphism $f^{-1} : B \rightarrow A$ which is a two-sided inverse of f in \mathcal{K} . The relation of isomorphism is reflexive, symmetric, and transitive, i.e., equivalence relation.

For example, an isomorphism in the category of sets is called a set-isomorphism, or a *bijection*, in the category of topological spaces is called a topological isomorphism, or a *homeomorphism*, in the category of differentiable manifolds is called a differentiable isomorphism, or a *diffeomorphism*.

A morphism $f \in \text{Mor}_{\mathcal{K}}(A, B)$ is *regular* if there exists a morphism $g : B \rightarrow A$ in \mathcal{K} such that $f \circ g \circ f = f$. Any morphism with either a left or a right inverse is regular.

An object T is a *terminal object* in \mathcal{K} if to each object $A \in \text{Ob}(\mathcal{K})$ there is exactly one arrow $A \rightarrow T$. An object S is an *initial object* in \mathcal{K} if to each object $A \in \text{Ob}(\mathcal{K})$ there is exactly one arrow $S \rightarrow A$. A *null object* $Z \in \text{Ob}(\mathcal{K})$ is an object which is both initial and terminal; it is unique up to isomorphism. For any two objects $A, B \in \text{Ob}(\mathcal{K})$ there is a unique morphism $A \rightarrow Z \rightarrow B$ (the composite through Z), called the *zero morphism* from A to B .

A notion of subcategory is analogous to the notion of subset. A subcategory \mathcal{L} of a category \mathcal{K} is said to be a *complete subcategory* iff for any objects $A, B \in \mathcal{L}$, every morphism $A \rightarrow B$ of \mathcal{L} is in \mathcal{K} .

A *groupoid* is a category in which every morphism is invertible. A typical groupoid is the *fundamental groupoid* $\Pi_1(X)$ of a topological space X . An object of $\Pi_1(X)$ is a point $x \in X$, and a morphism $x \rightarrow x'$ of $\Pi_1(X)$ is a homotopy class of paths f from x to x' . The *composition* of paths $g : x' \rightarrow x''$ and $f : x \rightarrow x'$ is the path h which is ‘ f followed by g ’. Composition applies also to homotopy classes, and makes $\Pi_1(X)$ a category and a groupoid (the inverse of any path is the same path traced in the opposite direction).

A *group* is a groupoid with one object, i.e., a *category with one object* in which *all morphisms are isomorphisms*. Therefore, if we try to generalize the concept of a group, keeping associativity as an essential property, we get the notion of a category.

A category is *discrete* if every morphism is an identity. A *monoid* is a category with one object. A *group* is a category with one object in which every morphism has a two-sided inverse under composition.

Homological algebra was the progenitor of category theory (see e.g., [Dieudonne (1988)]). Generalizing L. Euler’s formula $f + v = e + 2$ for the faces, vertices and edges of a convex polyhedron, E. Betti defined *numerical*

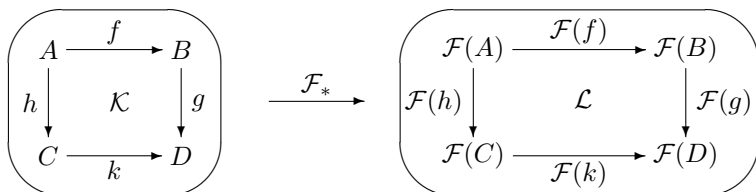
invariants of spaces by formal addition and subtraction of faces of various dimensions; H. Poincaré formalized these and introduced homology. E. Noether stressed the fact that these calculations go on in Abelian groups, and that the operation ∂_n taking a face of dimension n to the alternating sum of faces of dimension $n-1$ which form its boundary is a homomorphism, and it also satisfies $\partial_n \circ \partial_{n+1} = 0$. There are many ways of approximating a given space by polyhedra, but the quotient $H_n = \text{Ker } \partial_n / \text{Im } \partial_{n+1}$ is an invariant, the *homology group*. Since Noether, the groups have been the object of study instead of their dimensions, which are the *Betti numbers*.

2.3.3 Functors

In algebraic topology, one attempts to assign to every topological space X some algebraic object $\mathcal{F}(X)$ in such a way that to every C^0 -function $f : X \rightarrow Y$ there is assigned a homomorphism $\mathcal{F}(f) : \mathcal{F}(X) \rightarrow \mathcal{F}(Y)$ (see [Switzer (1975); Dodson and Parker (1997)]). One advantage of this procedure is, e.g., that if one is trying to prove the non-existence of a C^0 -function $f : X \rightarrow Y$ with certain properties, one may find it relatively easy to prove the non-existence of the corresponding algebraic function $\mathcal{F}(f)$ and hence deduce that f could not exist. In other words, \mathcal{F} is to be a ‘homomorphism’ from one category (e.g., \mathcal{T}) to another (e.g., \mathcal{G} or \mathcal{A}). Formalization of this notion is a *functor*.

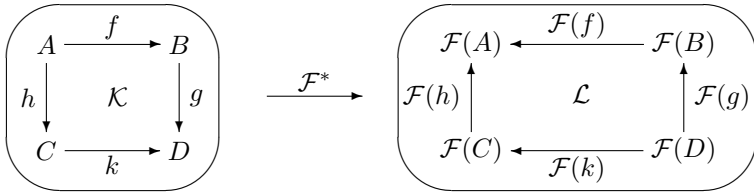
A functor is a generic *picture* projecting one category into another. Let $\mathcal{K} = (\text{Ob}(\mathcal{K}), \text{Mor}(\mathcal{K}))$ be a *source* (or domain) *category* and $\mathcal{L} = (\text{Ob}(\mathcal{L}), \text{Mor}(\mathcal{L}))$ be a *target* (or codomain) category. A functor $\mathcal{F} = (\mathcal{F}_O, \mathcal{F}_M)$ is defined as a pair of maps, $\mathcal{F}_O : \text{Ob}(\mathcal{K}) \rightarrow \text{Ob}(\mathcal{L})$ and $\mathcal{F}_M : \text{Mor}(\mathcal{K}) \rightarrow \text{Mor}(\mathcal{L})$, preserving categorical symmetry (i.e., commutativity of all diagrams) of \mathcal{K} in \mathcal{L} .

More precisely, a *covariant functor*, or simply a *functor*, $\mathcal{F}_* : \mathcal{K} \rightarrow \mathcal{L}$ is a *picture* in the target category \mathcal{L} of (all objects and morphisms of) the source category \mathcal{K} :



Similarly, a *contravariant functor*, or a *cofunctor*, $\mathcal{F}^* : \mathcal{K} \rightarrow \mathcal{L}$ is a *dual*

picture with reversed arrows:



In other words, a *functor* $\mathcal{F} : \mathcal{K} \rightarrow \mathcal{L}$ from a *source* category \mathcal{K} to a *target* category \mathcal{L} , is a pair $\mathcal{F} = (\mathcal{F}_O, \mathcal{F}_M)$ of maps $\mathcal{F}_O : \mathbf{Ob}(\mathcal{K}) \rightarrow \mathbf{Ob}(\mathcal{L})$, $\mathcal{F}_M : \mathbf{Mor}(\mathcal{K}) \rightarrow \mathbf{Mor}(\mathcal{L})$, such that

- (1) If $f \in \mathbf{Mor}_{\mathcal{K}}(A, B)$ then $\mathcal{F}_M(f) \in \mathbf{Mor}_{\mathcal{L}}(\mathcal{F}_O(A), \mathcal{F}_O(B))$ in case of the *covariant* functor \mathcal{F}_* , and $\mathcal{F}_M(f) \in \mathbf{Mor}_{\mathcal{L}}(\mathcal{F}_O(B), \mathcal{F}_O(A))$ in case of the *contravariant* functor \mathcal{F}^* ;
- (2) For all $A \in \mathbf{Ob}(\mathcal{K})$: $\mathcal{F}_M(1_A) = 1_{\mathcal{F}_O(A)}$;
- (3) For all $f, g \in \mathbf{Mor}(\mathcal{K})$: if $\text{cod}(f) = \text{dom}(g)$, then
 $\mathcal{F}_M(g \circ f) = \mathcal{F}_M(g) \circ \mathcal{F}_M(f)$ in case of the *covariant* functor \mathcal{F}_* , and
 $\mathcal{F}_M(g \circ f) = \mathcal{F}_M(f) \circ \mathcal{F}_M(g)$ in case of the *contravariant* functor \mathcal{F}^* .

Category theory originated in algebraic topology, which tried to assign algebraic invariants to topological structures. The golden rule of such *invariants* is that they should be *functors*. For example, the *fundamental group* π_1 is a functor. Algebraic topology constructs a group called the *fundamental group* $\pi_1(X)$ from any topological space X , which keeps track of how many holes the space X has. But also, any map between topological spaces determines a homomorphism $\phi : \pi_1(X) \rightarrow \pi_1(Y)$ of the fundamental groups. So the fundamental group is really a functor $\pi_1 : \mathcal{T} \rightarrow \mathcal{G}$. This allows us to completely transpose any situation involving *spaces* and *continuous maps* between them to a parallel situation involving *groups* and *homomorphisms* between them, and thus reduce some topology problems to algebra problems.

Also, singular homology in a given dimension n assigns to each topological space X an Abelian group $H_n(X)$, its *nth homology group* of X , and also to each continuous map $f : X \rightarrow Y$ of spaces a corresponding homomorphism $H_n(f) : H_n(X) \rightarrow H_n(Y)$ of groups, and this in such a way that $H_n(X)$ becomes a functor $H_n : \mathcal{T} \rightarrow \mathcal{A}$.

The leading idea in the *use of functors in topology* is that H_n or π_n gives an algebraic picture or image not just of the topological spaces X, Y but also of all the continuous maps $f : X \rightarrow Y$ between them.

Similarly, there is a functor $\Pi_1 : \mathcal{T} \rightarrow \mathcal{G}$, called the ‘fundamental groupoid functor’, which plays a very basic role in algebraic topology. Here’s how we get from any space X its ‘fundamental groupoid’ $\Pi_1(X)$. To say what the groupoid $\Pi_1(X)$ is, we need to say what its objects and morphisms are. The objects in $\Pi_1(X)$ are just the *points* of X and the morphisms are just certain equivalence classes of *paths* in X . More precisely, a morphism $f : x \rightarrow y$ in $\Pi_1(X)$ is just an equivalence class of continuous paths from x to y , where two paths from x to y are decreed equivalent if one can be continuously deformed to the other while not moving the endpoints. (If this equivalence relation holds we say the two paths are ‘homotopic’, and we call the equivalence classes ‘homotopy classes of paths’ (see [MacLane (1971); Switzer (1975)]).

Another examples are covariant *forgetful* functors:

- From the category of topological spaces to the category of sets; it ‘forgets’ the topology–structure.
- From the category of metric spaces to the category of topological spaces with the topology induced by the metrics; it ‘forgets’ the metric.

For each category \mathcal{K} , the *identity functor* $I_{\mathcal{K}}$ takes every \mathcal{K} –object and every \mathcal{K} –morphism to itself.

Given a category \mathcal{K} and its subcategory \mathcal{L} , we have an *inclusion functor* $\text{In} : \mathcal{K} \rightarrow \mathcal{K}$.

Given a category \mathcal{K} , a *diagonal functor* $\Delta : \mathcal{K} \rightarrow \mathcal{K}$ takes each object $A \in \mathcal{K}$ to the object (A, A) in the product category $\mathcal{K} \times \mathcal{K}$.

Given a category \mathcal{K} and a category of sets \mathcal{S} , each object $A \in \mathcal{K}$ determines a *covariant Hom–functor* $\mathcal{K}[A, _] : \mathcal{K} \rightarrow \mathcal{S}$, a *contravariant Hom–functor* $\mathcal{K}[_, A] : \mathcal{K} \rightarrow \mathcal{S}$, and a *Hom–bifunctor* $\mathcal{K}[_, _] : \mathcal{K}^{op} \times \mathcal{K} \rightarrow \mathcal{S}$.

A functor $\mathcal{F} : \mathcal{K} \rightarrow \mathcal{L}$ is a *faithful functor* if for all $A, B \in \text{Ob}(\mathcal{K})$ and for all $f, g \in \text{Mor}_{\mathcal{K}}(A, B)$, $\mathcal{F}(f) = \mathcal{F}(g)$ implies $f = g$; it is a *full functor* if for every $h \in \text{Mor}_{\mathcal{L}}(\mathcal{F}(A), \mathcal{F}(B))$, there is $g \in \text{Mor}_{\mathcal{K}}(A, B)$ such that $h = \mathcal{F}(g)$; it is a *full embedding* if it is both full and faithful.

A *representation of a group* is a functor $\mathcal{F} : \mathcal{G} \rightarrow \mathcal{V}$.

Similarly, we can define a *representation of a category* to be a functor $\mathcal{F} : \mathcal{K} \rightarrow \mathcal{V}$ from the 2–category \mathcal{K} (a ‘big’ category including all ordinary, or ‘small’ categories, see section (2.3.8.1) below) to the category of vector spaces \mathcal{V} . In this way, a category is a generalization of a group and group representations are a special case of category representations.

2.3.4 Natural Transformations

A *natural transformation* (i.e., a *functor morphism*) $\tau : \mathcal{F} \rightarrow \mathcal{G}$ is a map between two functors of the same variance, $(\mathcal{F}, \mathcal{G}) : \mathcal{K} \Rightarrow \mathcal{L}$, preserving categorical symmetry:

$$\begin{array}{ccc}
 \boxed{\begin{array}{c} A \xrightarrow{f} B \\ \mathcal{K} \end{array}} & \begin{array}{c} \xrightarrow{\mathcal{F}} \\ \tau \Downarrow \\ \xrightarrow{\mathcal{G}} \end{array} & \boxed{\begin{array}{ccc} \mathcal{F}(A) & \xrightarrow{\mathcal{F}(f)} & \mathcal{F}(B) \\ \tau_A \downarrow & \mathcal{L} & \downarrow \tau_B \\ \mathcal{G}(A) & \xrightarrow{\mathcal{G}(f)} & \mathcal{G}(B) \end{array}}
 \end{array}$$

More precisely, all functors of the same variance from a source category \mathcal{K} to a target category \mathcal{L} form themselves objects of the *functor category* $\mathcal{L}^{\mathcal{K}}$. Morphisms of $\mathcal{L}^{\mathcal{K}}$, called *natural transformations*, are defined as follows.

Let $\mathcal{F} : \mathcal{K} \rightarrow \mathcal{L}$ and $\mathcal{G} : \mathcal{K} \rightarrow \mathcal{L}$ be two functors of the same variance from a category \mathcal{K} to a category \mathcal{L} . Natural transformation $\mathcal{F} \xrightarrow{\tau} \mathcal{G}$ is a family of morphisms such that for all $f \in \text{Mor}_{\mathcal{K}}(A, B)$ in the source category \mathcal{K} , we have $\mathcal{G}(f) \circ \tau_A = \tau_B \circ \mathcal{F}(f)$ in the target category \mathcal{L} . Then we say that the *component* $\tau_A : \mathcal{F}(A) \rightarrow \mathcal{G}(A)$ is *natural in A*.

If we think of a functor \mathcal{F} as giving a *picture* in the target category \mathcal{L} of (all the objects and morphisms of) the source category \mathcal{K} , then a natural transformation τ represents a set of morphisms mapping the picture \mathcal{F} to another picture \mathcal{G} , preserving the commutativity of all diagrams.

An invertible natural transformation, such that all components τ_A are isomorphisms) is called a *natural equivalence* (or, *natural isomorphism*). In this case, the inverses $(\tau_A)^{-1}$ in \mathcal{L} are the components of a natural isomorphism $(\tau)^{-1} : \mathcal{G} \xrightarrow{*} \mathcal{F}$. Natural equivalences are among the most important *metamathematical constructions* in algebraic topology (see [Switzer (1975)]).

For example, let \mathcal{B} be the category of Banach spaces over \mathbb{R} and bounded linear maps. Define $D : \mathcal{B} \rightarrow \mathcal{B}$ by taking $D(X) = X^* = \text{Banach space of bounded linear functionals on a space } X$ and $D(f) = f^*$ for $f : X \rightarrow Y$ a bounded linear map. Then D is a cofunctor. $D^2 = D \circ D$ is also a functor. We also have the identity functor $1 : \mathcal{B} \rightarrow \mathcal{B}$. Define $T : 1 \rightarrow D \circ D$ as follows: for every $X \in \mathcal{B}$ let $T(X) : X \rightarrow D^2X = X^{**}$ be the *natural inclusion* – that is, for $x \in X$ we have $[T(X)(x)](f) = f(x)$ for every $f \in X^*$. T is a natural transformation. On the subcategory of nD Banach spaces T is even a natural equivalence. The largest subcategory of \mathcal{B} on which T is a natural equivalence is called the category of reflexive Banach

spaces [Switzer (1975)].

As S. Eilenberg and S. MacLane first observed, ‘category’ has been defined in order to define ‘functor’ and ‘functor’ has been defined in order to define ‘natural transformation’ [MacLane (1971)].

2.3.4.1 Compositions of Natural Transformations

Natural transformations can be *composed* in two different ways. First, we have an ‘ordinary’ composition: if \mathcal{F}, \mathcal{G} and \mathcal{H} are three functors from the source category \mathcal{A} to the target category \mathcal{B} , and then $\alpha : \mathcal{F} \rightarrow \mathcal{G}, \beta : \mathcal{G} \rightarrow \mathcal{H}$ are two natural transformations, then the formula

$$(\beta \circ \alpha)_A = \beta_A \circ \alpha_A, \quad \text{for all } A \in \mathcal{A}, \quad (2.61)$$

defines a new natural transformation $\beta \circ \alpha : \mathcal{F} \rightarrow \mathcal{H}$. This composition law is clearly associative and possesses a unit $1_{\mathcal{F}}$ at each functor \mathcal{F} , whose \mathcal{A} –component is $1_{\mathcal{F}\mathcal{A}}$.

Second, we have the *Godement product* of natural transformations, usually denoted by $*$. Let \mathcal{A}, \mathcal{B} and \mathcal{C} be three categories, $\mathcal{F}, \mathcal{G}, \mathcal{H}$ and \mathcal{K} be four functors such that $(\mathcal{F}, \mathcal{G}) : \mathcal{A} \rightrightarrows \mathcal{B}$ and $(\mathcal{H}, \mathcal{K}) : \mathcal{B} \rightrightarrows \mathcal{C}$, and $\alpha : \mathcal{F} \rightarrow \mathcal{G}, \beta : \mathcal{H} \rightarrow \mathcal{K}$ be two natural transformations. Now, instead of (2.61), the Godement composition is given by

$$(\beta * \alpha)_A = \beta_{GA} \circ H(\alpha_A) = K(\alpha_A) \circ \beta_{FA}, \quad \text{for all } A \in \mathcal{A}, \quad (2.62)$$

which defines a new natural transformation $\beta * \alpha : \mathcal{H} \circ \mathcal{F} \rightarrow \mathcal{K} \circ \mathcal{G}$.

Finally, the two compositions (2.61) and (2.61) of natural transformations can be combined as

$$(\delta * \gamma) \circ (\beta * \alpha) = (\delta \circ \beta) * (\gamma \circ \alpha),$$

where \mathcal{A}, \mathcal{B} and \mathcal{C} are three categories, $\mathcal{F}, \mathcal{G}, \mathcal{H}, \mathcal{K}, \mathcal{L}, \mathcal{M}$ are six functors, and $\alpha : \mathcal{F} \rightarrow \mathcal{H}, \beta : \mathcal{G} \rightarrow \mathcal{K}, \gamma : \mathcal{H} \rightarrow \mathcal{L}, \delta : \mathcal{K} \rightarrow \mathcal{M}$ are four natural transformations.

2.3.4.2 Dinatural Transformations

Double natural transformations are called *dinatural transformations*. An *end of a functor* $S : C^{op} \times C \rightarrow X$ is a universal dinatural transformation from a constant e to S . In other words, an end of S is a pair $\langle e, \omega \rangle$, where e is an object of X and $\omega : e \rightarrow S$ is a *wedge (dinatural) transformation* with the property that to every wedge $\beta : x \rightarrow S$ there is a unique arrow

$h : x \rightarrow e$ of B with $\beta_c = \omega_c h$ for all $a \in C$. We call ω the *ending wedge* with *components* ω_c , while the object e itself, by abuse of language, is called the end of S and written with integral notation as $\int_c S(c, c)$; thus

$$S(c, c) \xrightarrow{\omega_c} \int_c S(c, c) = e.$$

Note that the ‘variable of integration’ c appears twice under the integral sign (once contravariant, once covariant) and is ‘bound’ by the integral sign, in that the result no longer depends on c and so is unchanged if ‘ c ’ is replaced by any other letter standing for an object of the category C . These properties are like those of the letter x under the usual integral symbol $\int f(x) dx$ of calculus.

Every end is manifestly a limit (see below) – specifically, a limit of a suitable diagram in X made up of pieces like $S(b, b) \rightarrow S(b, c) \rightarrow S(c, c)$.

For each functor $T : C \rightarrow X$ there is an isomorphism

$$\int_c S(c, c) = \int_c Tc \cong \text{Lim } T,$$

valid when either the end of the limit exists, carrying the ending wedge to the limiting cone; the indicated notation thus allows us to write any limit as an integral (an end) without explicitly mentioning the dummy variable (the first variable c of S).

A functor $H : X \rightarrow Y$ is said to *preserve the end* of a functor $S : C^{op} \times C \rightarrow X$ when $\omega : e \rightrightarrows S$ an end of S in X implies that $H\omega : He \rightrightarrows HS$ is an and for HS ; in symbols

$$H \int_c S(c, c) = \int_c HS(c, c).$$

Similarly, H *creates* the end of S when to each end $v : y \rightrightarrows HS$ in Y there is a unique wedge $\omega : e \rightrightarrows S$ with $H\omega = v$, and this wedge ω is an end of S .

The definition of the coend of a functor $S : C^{op} \times C \rightarrow X$ is dual to that of an end. A *coend* of S is a pair $\langle d, \zeta \rangle$, consisting of an object $d \in X$ and a wedge $\zeta : S \rightrightarrows d$. The object d (when it exists, unique up to isomorphism) will usually be written with an integral sign and with the bound variable c as superscript; thus

$$S(c, c) \xrightarrow{\zeta_c} \int_c S(c, c) = d.$$

The formal properties of coends are dual to those of ends. Both are much like those for integrals in calculus (see [MacLane (1971)], for technical details).

2.3.5 Limits and Colimits

In abstract algebra constructions are often defined by an abstract property which requires the existence of unique morphisms under certain conditions. These properties are called *universal properties*. The *limit* of a functor generalizes the notions of inverse limit and product used in various parts of mathematics. The dual notion, *colimit*, generalizes direct limits and direct sums. Limits and colimits are defined via universal properties and provide many examples of *adjoint functors*.

A *limit* of a covariant functor $\mathcal{F} : \mathcal{J} \rightarrow \mathcal{C}$ is an object L of \mathcal{C} , together with morphisms $\phi_X : L \rightarrow \mathcal{F}(X)$ for every object X of \mathcal{J} , such that for every morphism $f : X \rightarrow Y$ in \mathcal{J} , we have $\mathcal{F}(f)\phi_X = \phi_Y$, and such that the following *universal property* is satisfied: for any object N of \mathcal{C} and any set of morphisms $\psi_X : N \rightarrow \mathcal{F}(X)$ such that for every morphism $f : X \rightarrow Y$ in \mathcal{J} , we have $\mathcal{F}(f)\psi_X = \psi_Y$, there exists precisely one morphism $u : N \rightarrow L$ such that $\phi_X u = \psi_X$ for all X . If \mathcal{F} has a limit (which it need not), then the limit is defined up to a unique isomorphism, and is denoted by $\lim \mathcal{F}$.

Analogously, a *colimit* of the functor $\mathcal{F} : \mathcal{J} \rightarrow \mathcal{C}$ is an object L of \mathcal{C} , together with morphisms $\phi_X : \mathcal{F}(X) \rightarrow L$ for every object X of \mathcal{J} , such that for every morphism $f : X \rightarrow Y$ in \mathcal{J} , we have $\phi_Y \mathcal{F}(f) = \phi_X$, and such that the following universal property is satisfied: for any object N of \mathcal{C} and any set of morphisms $\psi_X : \mathcal{F}(X) \rightarrow N$ such that for every morphism $f : X \rightarrow Y$ in \mathcal{J} , we have $\psi_Y \mathcal{F}(f) = \psi_X$, there exists precisely one morphism $u : L \rightarrow N$ such that $u\phi_X = \psi_X$ for all X . The colimit of \mathcal{F} , unique up to unique isomorphism if it exists, is denoted by $\text{colim } \mathcal{F}$.

Limits and colimits are related as follows: A functor $\mathcal{F} : \mathcal{J} \rightarrow \mathcal{C}$ has a colimit iff for every object N of \mathcal{C} , the functor $X \mapsto \text{Mor}_{\mathcal{C}}(\mathcal{F}(X), N)$ (which is a covariant functor on the dual category \mathcal{J}^{op}) has a limit. If that is the case, then $\text{Mor}_{\mathcal{C}}(\text{colim } \mathcal{F}, N) = \lim \text{Mor}_{\mathcal{C}}(\mathcal{F}(-), N)$ for every object N of \mathcal{C} .

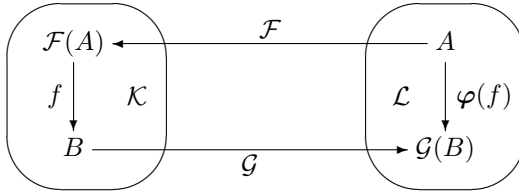
2.3.6 Adjunction

The most important functorial operation is adjunction; as S. MacLane once said, “Adjoint functors arise everywhere” [MacLane (1971)].

The *adjunction* $\varphi : \mathcal{F} \dashv \mathcal{G}$ between two functors $(\mathcal{F}, \mathcal{G}) : \mathcal{K} \rightleftharpoons \mathcal{L}$ of *opposite variance* [Kan (1958)], represents a *weak functorial inverse*

$$\frac{f : \mathcal{F}(A) \rightarrow B}{\varphi(f) : A \rightarrow \mathcal{G}(B)}$$

forming a *natural equivalence* $\varphi : \text{Mor}_{\mathcal{K}}(\mathcal{F}(A), B) \xrightarrow{\varphi} \text{Mor}_{\mathcal{L}}(A, \mathcal{G}(B))$. The adjunction isomorphism is given by a *bijective correspondence* (a 1–1 and onto map on objects) $\varphi : \text{Mor}(\mathcal{K}) \ni f \rightarrow \varphi(f) \in \text{Mor}(\mathcal{L})$ of isomorphisms in the two categories, \mathcal{K} (with a representative object A), and \mathcal{L} (with a representative object B). It can be depicted as a (non-commutative) diagram



In this case \mathcal{F} is called *left adjoint*, while \mathcal{G} is called *right adjoint*.

In other words, an *adjunction* $\mathcal{F} \dashv \mathcal{G}$ between two functors $(\mathcal{F}, \mathcal{G})$ of opposite variance, from a source category \mathcal{K} to a target category \mathcal{L} , is denoted by $(\mathcal{F}, \mathcal{G}, \eta, \varepsilon) : \mathcal{K} \rightleftharpoons \mathcal{L}$. Here, $\mathcal{F} : \mathcal{L} \rightarrow \mathcal{K}$ is the *left (upper) adjoint functor*, $\mathcal{G} : \mathcal{K} \rightarrow \mathcal{L}$ is the *right (lower) adjoint functor*, $\eta : 1_{\mathcal{L}} \rightarrow \mathcal{G} \circ \mathcal{F}$ is the *unit natural transformation* (or, *front adjunction*), and $\varepsilon : \mathcal{F} \circ \mathcal{G} \rightarrow 1_{\mathcal{K}}$ is the *counit natural transformation* (or, *back adjunction*).

For example, $\mathcal{K} = \mathcal{S}$ is the category of sets and $\mathcal{L} = \mathcal{G}$ is the category of groups. Then \mathcal{F} turns any set into the *free group* on that set, while the ‘forgetful’ functor \mathcal{F}^* turns any group into the *underlying set* of that group. Similarly, all sorts of other ‘free’ and ‘underlying’ constructions are also left and right adjoints, respectively.

Right adjoints preserve *limits*, and left adjoints preserve *colimits*.

The category \mathcal{C} is called a *cocomplete category* if every functor $\mathcal{F} : \mathcal{J} \rightarrow \mathcal{C}$ has a colimit. The following categories are cocomplete: $\mathcal{S}, \mathcal{G}, \mathcal{A}, \mathcal{T}$, and \mathcal{PT} .

The importance of adjoint functors lies in the fact that every functor which has a left adjoint (and therefore is a right adjoint) is continuous. In the category \mathcal{A} of Abelian groups, this e.g., shows that the kernel of a product of homomorphisms is naturally identified with the product of the kernels. Also, limit functors themselves are continuous. A covariant functor

$\mathcal{F} : \mathcal{J} \rightarrow \mathcal{C}$ is *cocontinuous* if it transforms colimits into colimits. Every functor which has a right adjoint (and is a left adjoint) is cocontinuous.

The *analogy* between *adjoint functors* and *adjoint linear operators* relies upon a deeper analogy: just as in quantum theory the inner product $\langle \phi, \psi \rangle$ represents the *amplitude* to pass from ϕ to ψ , in category theory $\text{Mor}(A, B)$ represents the *set of ways* to go from A to B . These are to Hilbert spaces as categories are to sets. The analogues of adjoint linear operators between Hilbert spaces are certain adjoint functors between 2-Hilbert spaces [Baez (1997); Baez and Dolan (1998)]. Similarly, the *adjoint representation* of a Lie group G is the linearized version of the action of G on itself by conjugation, i.e., for each $g \in G$, the inner automorphism $x \mapsto gxg^{-1}$ gives a linear transformation $\text{Ad}(g) : \mathfrak{g} \rightarrow \mathfrak{g}$, from the Lie algebra \mathfrak{g} of G to itself.

2.3.7 Abelian Categorical Algebra

An *Abelian category* is a certain kind of category in which morphisms and objects can be added and in which kernels and cokernels exist and have the usual properties. The motivating prototype example of an Abelian category is the category of Abelian groups \mathcal{A} . Abelian categories are the framework for homological algebra (see [Dieudonne (1988)]).

Given a homomorphism $f : A \rightarrow B$ between two objects $A \equiv \text{Dom } f$ and $B \equiv \text{Cod } f$ in an Abelian category \mathcal{A} , then its *kernel*, *image*, *cokernel* and *coimage* in \mathcal{A} are defined respectively as:

$$\begin{aligned} \text{Ker } f &= f^{-1}(e_B), & \text{Coker } f &= \text{Cod } f / \text{Im } f, \\ \text{Im } f &= f(A), & \text{Coim } f &= \text{Dom } f / \text{Ker } f, \end{aligned}$$

where e_B is a unit of B [Dodson and Parker (1997)].

In an Abelian category \mathcal{A} a *composable* pair of arrows,

$$\bullet \xrightarrow{f} B \xrightarrow{g} \bullet$$

is *exact* at B iff $\text{Im } f \equiv \text{Ker } g$ (equivalence as subobjects of B) – or, equivalently, if $\text{Coker } f \equiv \text{Coim } g$ [MacLane (1971)].

For each arrow f in an Abelian category \mathcal{A} the *triangular identities* read

$$\text{Ker}(\text{Coker}(\text{Ker } f)) = \text{Ker } f, \quad \text{Coker}(\text{Ker}(\text{Coker } f)) = \text{Coker } f.$$

The diagram (with 0 the null object)

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0 \quad (2.63)$$

is a *short exact sequence* when it is exact at A , at B , and at C .

Since $0 \rightarrow a$ is the zero arrow, exactness at A means just that f is *monic* (i.e., 1-1, or injective map); dually, exactness at C means that g is *epic* (i.e., onto, or surjective map). Therefore, (2.63) is equivalent to

$$f = \text{Ker } g, \quad g = \text{Coker } f.$$

Similarly, the statement that $h = \text{Coker } f$ becomes the statement that the sequence

$$A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

is exact at B and at C . Classically, such a sequence was called a short right exact sequence. Similarly, $k = \text{Ker } f$ is expressed by a short left exact sequence

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C.$$

If \mathcal{A} and \mathcal{A}' are Abelian categories, an *additive functor* $\mathcal{F} : \mathcal{A} \rightarrow \mathcal{A}'$ is a functor from \mathcal{A} to \mathcal{A}' with

$$\mathcal{F}(f + f') = \mathcal{F}f + \mathcal{F}f',$$

for any parallel pair of arrows $f, f' : b \rightarrow c$ in \mathcal{A} . It follows that $\mathcal{F}0 = 0$.

A functor $\mathcal{F} : \mathcal{A} \rightarrow \mathcal{A}'$ between Abelian categories \mathcal{A} and \mathcal{A}' is, by definition, *exact* when it preserves all finite limits and all finite colimits. In particular, an exact functor preserves kernels and cokernels, which means that

$$\text{Ker}(\mathcal{F}f) = \mathcal{F}(\text{Ker } f) \quad \text{and} \quad \text{Coker}(\mathcal{F}f) = \mathcal{F}(\text{Coker } f);$$

then \mathcal{F} also preserves images, coimages, and carries exact sequences to exact sequences. By construction of limits from products and equalizers and dual constructions, $\mathcal{F} : \mathcal{A} \rightarrow \mathcal{A}'$ is exact iff it is additive and preserves kernels and cokernels.

A functor \mathcal{F} is *left exact* when it preserves all finite limits. In other words, \mathcal{F} is left exact iff it is additive and $\text{Ker}(\mathcal{F}f) = \mathcal{F}(\text{Ker } f)$ for all f : the last condition is equivalent to the requirement that \mathcal{F} preserves short left exact sequences.

Similarly, a functor \mathcal{F} is *right exact* when it preserves all finite colimits. In other words, \mathcal{F} is right exact iff it is additive and $\text{Coker}(\mathcal{F}f) =$

$\mathcal{F}(\text{Coker } f)$ for all f : the last condition is equivalent to the requirement that \mathcal{F} preserves short right exact sequences.

In an Abelian category \mathcal{A} , a *chain complex* is a sequence

$$\dots \longrightarrow c_{n+1} \xrightarrow{\partial_{n+1}} c_n \xrightarrow{\partial_n} c_{n-1} \longrightarrow \dots$$

of composable arrows, with $\partial_n \partial_{n+1} = 0$ for all n . The sequence need not be exact at c_n ; the deviation from exactness is measured by the n th *homology object*

$$H_n c = \text{Ker}(\partial_n : c_n \longrightarrow c_{n-1}) / \text{Im}(\partial_{n+1} : c_{n+1} \longrightarrow c_n).$$

Similarly, a *cochain complex* in an Abelian category \mathcal{A} is a sequence

$$\dots \longrightarrow w_{n+1} \xrightarrow{d_{n+1}} w_n \xrightarrow{d_n} w_{n-1} \longrightarrow \dots$$

of composable arrows, with $d_n d_{n+1} = 0$ for all n . The sequence need not be exact at w_n ; the deviation from exactness is measured by the n th *cohomology object*

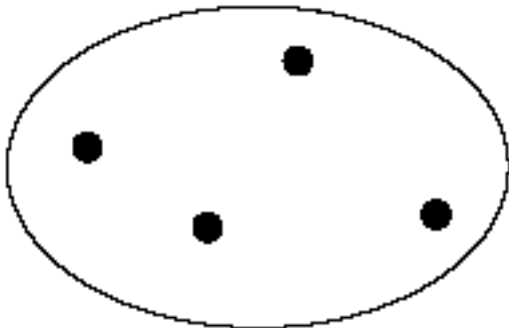
$$H^n w = \text{Ker}(d_{n+1} : w_n \longrightarrow w_{n+1}) / \text{Im}(d_n : w_{n-1} \longrightarrow w_n).$$

A *cycle* is a chain C such that $\partial C = 0$. A *boundary* is a chain C such that $C = \partial B$, for any other chain B .

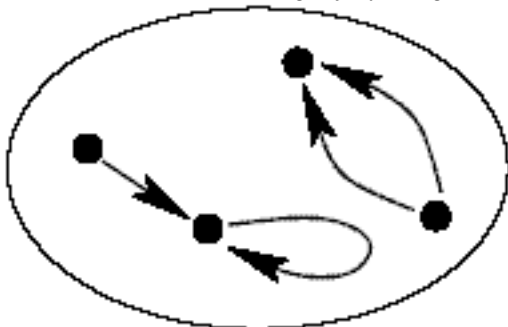
A *cocycle* (a *closed form*) is a cochain ω such that $d\omega = 0$. A *coboundary* (an *exact form*) is a cochain ω such that $\omega = d\theta$, for any other cochain θ .

2.3.8 n -Categories

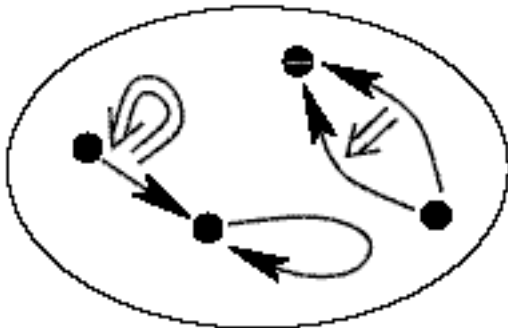
In this subsection we introduce the concept of modern n -categories. Intuitively, in describing dynamical systems (processes) by means of n -categories, instead of classical starting with a *set of things*:



we can now start with a *category of things and processes between things*:

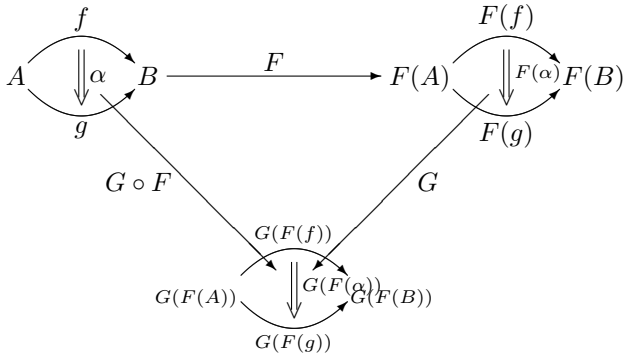


or, a *2-category of things, processes, and processes between processes*:

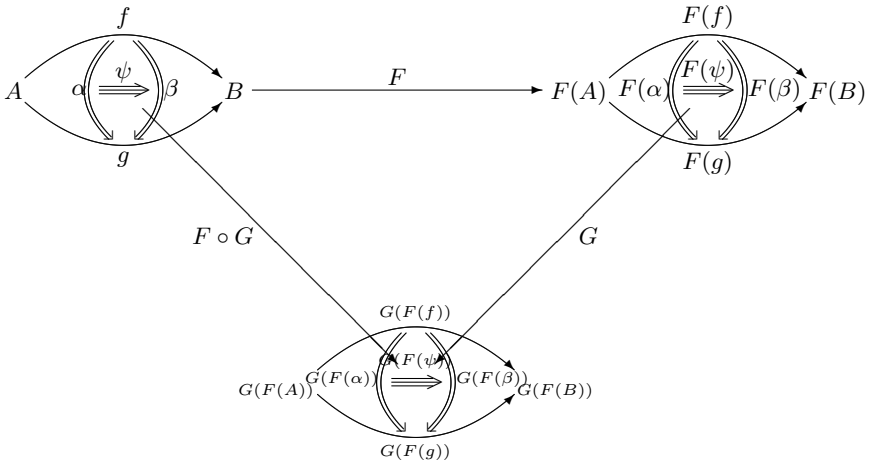


... and so on. For example, this n -categorical framework can be used for higher gauge theory [Baez (2002)], which .

a 2–category is a commutative triangle:



a 3–category is a commutative triangle:



etc., up to n –categories.

Many deep-sounding results in mathematical sciences are get by the process of *categorification*¹⁴ of the high school mathematics [Crane and Frenkel (1994); Baez and Dolan (1998)].

An n –category is a generic mathematical structure consisting of a collection of objects, a collection of arrows between objects, a collection of 2–arrows between arrows [Bénabou (1967)], a collection of 3–arrows be-

¹⁴Categorification means replacing sets with categories, functions with functors, and equations between functions by natural equivalences between functors. Iterating this process requires a theory of n –categories.

tween 2-arrows, and so on up to n [Baez (1997); Baez and Dolan (1998); Leinster (2002); Leinster (2003); Leinster (2004)].

More precisely, an n -category (for $n \geq 0$) consists of:

- 0-cells, or objects, A, B, \dots
- 1-cells, or arrows, $A \xrightarrow{f} B$, with a composition

$$A \xrightarrow{f} B \xrightarrow{g} C = A \xrightarrow{g \circ f} C$$

- 2-cells, ‘arrows between arrows’, $A \begin{array}{c} \xrightarrow{f} \\ \Downarrow \alpha \\ \xrightarrow{g} \end{array} B$, with vertical compositions

(denoted by \circ) and horizontal compositions (denoted by $*$), respectively given by

$$A \begin{array}{c} \xrightarrow{f} \\ \Downarrow \alpha \\ \xrightarrow{g} \\ \Downarrow \beta \\ \xrightarrow{h} \end{array} B = A \begin{array}{c} \xrightarrow{f} \\ \Downarrow \beta \circ \alpha \\ \xrightarrow{h} \end{array} B \quad \text{and}$$

$$A \begin{array}{c} \xrightarrow{f} \\ \Downarrow \alpha \\ \xrightarrow{g} \end{array} A' \begin{array}{c} \xrightarrow{f'} \\ \Downarrow \alpha' \\ \xrightarrow{g'} \end{array} A'' = A \begin{array}{c} \xrightarrow{f' \circ f} \\ \Downarrow \alpha' * \alpha \\ \xrightarrow{g' \circ g} \end{array} A''$$

- 3-cells, ‘arrows between arrows

$$\text{between arrows’, } A \begin{array}{c} \xrightarrow{f} \\ \alpha \left(\begin{array}{c} \xrightarrow{\Gamma} \\ \Downarrow \\ \xrightarrow{\beta} \end{array} \right) \xrightarrow{g} \end{array} B \quad (\text{where the } \Gamma\text{-arrow goes in}$$

a direction perpendicular to f and α), with various kinds of vertical, horizontal and mixed compositions,

- etc., up to n -cells.

Calculus of n -categories has been developed as follows. First, there is \mathcal{K}_2 , the 2-category of all ordinary (or small) categories. \mathcal{K}_2 has cate-

gories $\mathcal{K}, \mathcal{L}, \dots$ as objects, functors $\mathcal{F}, \mathcal{G} : \mathcal{K} \rightrightarrows \mathcal{L}$ as arrows, and natural transformations, like $\tau : \mathcal{F} \rightarrow \mathcal{G}$ as 2-arrows.

In a similar way, the arrows in a 3-category \mathcal{K}_3 are 2-functors $\mathcal{F}_2, \mathcal{G}_2, \dots$ sending objects in \mathcal{K}_2 to objects in \mathcal{L}_2 , arrows to arrows, and 2-arrows to 2-arrows, strictly preserving all the structure of \mathcal{K}_2

$$\begin{array}{ccc}
 \begin{array}{c} f \\ \curvearrowright \\ A \Downarrow \alpha B \\ \curvearrowleft \\ g \end{array} & \xrightarrow{\mathcal{F}_2} & \begin{array}{c} \mathcal{F}_2(f) \\ \curvearrowright \\ \mathcal{F}_2(A) \Downarrow \mathcal{F}_2(\alpha) \mathcal{F}_2(B) \\ \curvearrowleft \\ \mathcal{F}_2(g) \end{array}
 \end{array}$$

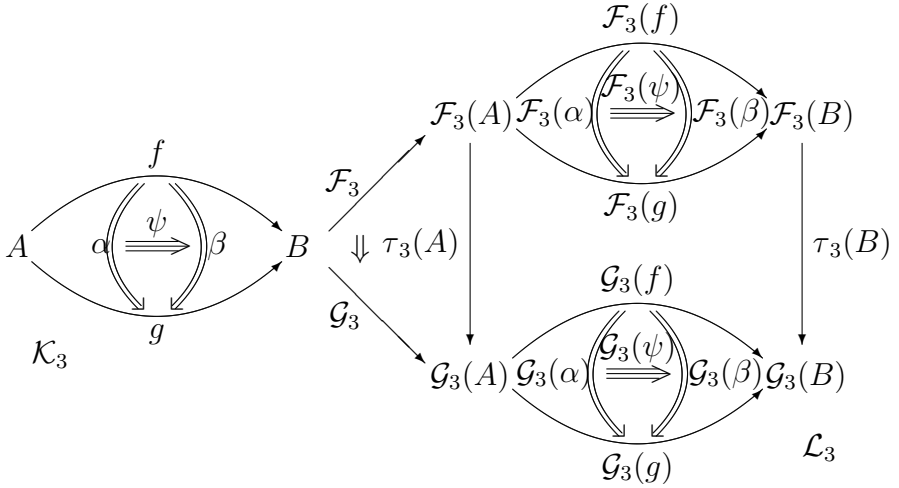
The 2-arrows in \mathcal{K}_3 are 2-natural transformations, like $\tau_2 : \mathcal{F}_2 \xRightarrow{2} \mathcal{G}_2$ between 2-functors $\mathcal{F}_2, \mathcal{G}_2 : \mathcal{K}_2 \rightarrow \mathcal{L}_2$ that sends each object in \mathcal{K}_2 to an arrow in \mathcal{L}_2 and each arrow in \mathcal{K}_2 to a 2-arrow in \mathcal{L}_2 , and satisfies natural transformation-like conditions. We can visualize τ_2 as a prism going from one functorial picture of \mathcal{K}_2 in \mathcal{L}_2 to another, built using commutative squares:

$$\begin{array}{ccc}
 \begin{array}{c} f \\ \curvearrowright \\ A \Downarrow \alpha B \\ \curvearrowleft \\ g \\ \mathcal{K}_2 \end{array} & \begin{array}{c} \nearrow \mathcal{F}_2 \\ \searrow \mathcal{G}_2 \end{array} & \begin{array}{c} \begin{array}{c} \mathcal{F}_2(f) \\ \curvearrowright \\ \mathcal{F}_2(A) \Downarrow \mathcal{F}_2(\alpha) \mathcal{F}_2(B) \\ \curvearrowleft \\ \mathcal{F}_2(g) \end{array} \\ \downarrow \tau_2(A) \\ \begin{array}{c} \mathcal{G}_2(f) \\ \curvearrowright \\ \mathcal{G}_2(A) \Downarrow \mathcal{G}_2(\alpha) \mathcal{G}_2(B) \\ \curvearrowleft \\ \mathcal{G}_2(g) \end{array} \end{array} \\
 & & \begin{array}{c} \downarrow \tau_2(B) \\ \mathcal{L}_2 \end{array}
 \end{array}$$

Similarly, the arrows in a 4-category \mathcal{K}_4 are 3-functors $\mathcal{F}_3, \mathcal{G}_3, \dots$ sending objects in \mathcal{K}_3 to objects in \mathcal{L}_3 , arrows to arrows, and 2-arrows to 2-arrows, strictly preserving all the structure of \mathcal{K}_3

$$\begin{array}{ccc}
 \begin{array}{c} f \\ \curvearrowright \\ A \Downarrow \alpha \Downarrow \psi \Downarrow \beta \\ \curvearrowleft \\ g \end{array} & \xrightarrow{\mathcal{F}_3} & \begin{array}{c} \mathcal{F}_3(f) \\ \curvearrowright \\ \mathcal{F}_3(A) \Downarrow \mathcal{F}_3(\alpha) \Downarrow \mathcal{F}_3(\psi) \Downarrow \mathcal{F}_3(\beta) \mathcal{F}_3(B) \\ \curvearrowleft \\ \mathcal{F}_3(g) \end{array}
 \end{array}$$

The 2-arrows in \mathcal{K}_4 are 3-natural transformations, like $\tau_3 : \mathcal{F} \xrightarrow{3} \mathcal{G}$ between 3-functors $\mathcal{F}_3, \mathcal{G}_3 : \mathcal{K}_3 \rightarrow \mathcal{L}_3$ that sends each object in \mathcal{K}_3 to a arrow in \mathcal{L}_3 and each arrow in \mathcal{K}_3 to a 2-arrow in \mathcal{L}_3 , and satisfies natural transformation-like conditions. We can visualize τ_3 as a prism going from one picture of \mathcal{K}_3 in \mathcal{L}_3 to another, built using commutative squares:



2.3.8.2 Topological Structure of n -Categories

We already emphasized the topological nature of ordinary category theory. This fact is even more obvious in the general case of n -categories (see [Leinster (2002); Leinster (2003); Leinster (2004)]).

2.3.8.3 Homotopy Theory and Related n -Categories

Any topological manifold M induces an n -category $\Pi_n(M)$ (its *fundamental n -groupoid*), in which 0-cells are *points* in M ; 1-cells are *paths* in M (i.e., parameterized continuous maps $f : [0, 1] \rightarrow M$); 2-cells are *homotopies* (denoted by \simeq) of *paths* relative to endpoints (i.e., parameterized continuous maps $h : [0, 1] \times [0, 1] \rightarrow M$); 3-cells are *homotopies of homotopies* of paths in M (i.e., parameterized continuous maps $j : [0, 1] \times [0, 1] \times [0, 1] \rightarrow M$); categorical *composition* is defined by *past-ing* paths and homotopies. In this way the following ‘homotopy skeleton’

emerges:

$$0 - \text{cell} : x \bullet \quad x \in M;$$

$$1 - \text{cell} : x \bullet \xrightarrow{f} \bullet y \quad f : x \simeq y \in M,$$

$$f : [0, 1] \rightarrow M, f : x \mapsto y, y = f(x), f(0) = x, f(1) = y;$$

$$\text{e.g., linear path: } f(t) = (1 - t)x + ty;$$

$$2 - \text{cell} : x \bullet \begin{array}{c} \xrightarrow{f} \\ \Downarrow h \\ \xrightarrow{g} \end{array} \bullet y \quad h : f \simeq g \in M,$$

$$h : [0, 1] \times [0, 1] \rightarrow M, h : f \mapsto g, g = h(f(x)),$$

$$h(x, 0) = f(x), h(x, 1) = g(x), h(0, t) = x, h(1, t) = y$$

$$\text{e.g., linear homotopy: } h(x, t) = (1 - t)f(x) + tg(x);$$

$$3 - \text{cell} : x \bullet \begin{array}{c} \xrightarrow{f} \\ \begin{array}{c} \curvearrowright \\ h \end{array} \begin{array}{c} \xrightarrow{j} \\ \rightleftarrows \\ i \end{array} \begin{array}{c} \curvearrowleft \\ g \end{array} \\ \xrightarrow{g} \end{array} \bullet y \quad j : h \simeq i \in M,$$

$$j : [0, 1] \times [0, 1] \times [0, 1] \rightarrow M, j : h \mapsto i, i = j(h(f(x)))$$

$$j(x, t, 0) = h(f(x)), j(x, t, 1) = i(f(x)),$$

$$j(x, 0, s) = f(x), j(x, 1, s) = g(x),$$

$$j(0, t, s) = x, j(1, t, s) = y$$

$$\text{e.g., linear composite homotopy: } j(x, t, s) = (1 - t)h(f(x)) + ti(f(x)).$$

If M is a *smooth* manifold, then all included paths and homotopies need to be *smooth*. Recall that a *groupoid* is a category in which every morphism is invertible; its special case with only one object is a *group*.

Category \mathcal{TT}

Topological n -category \mathcal{TT} has:

- 0-cells: topological spaces X
- 1-cells: continuous maps $X \xrightarrow{f} Y$

- 2-cells: homotopies h between f and $g : X \rightrightarrows Y$

i.e., continuous maps $h : X \times [0, 1] \rightarrow Y$, such that $\forall x \in X$, $h(x, 0) = f(x)$ and $h(x, 1) = g(x)$

- 3-cells: homotopies between homotopies : $X \rightrightarrows Y$

i.e., continuous maps $j : X \times [0, 1] \times [0, 1] \rightarrow Y$.

Category \mathcal{CK}

Consider an n -category \mathcal{CK} , which has:

- 0-cells: chain complexes A (of Abelian groups, say)
- 1-cells: chain maps $A \xrightarrow{f} B$

- 2-cells: chain homotopies $A \rightrightarrows B$,

i.e., maps $\alpha : A \rightarrow B$ of degree 1

- 3-cells $A \rightrightarrows B$: homotopies between homotopies,

i.e., maps $\Gamma : A \rightarrow B$ of degree 2 such that $d\Gamma - \Gamma d = \beta - \alpha$.

There ought to be some kind of map $\mathcal{CC} : \mathcal{TT} \Rightarrow \mathcal{CK}$ (see [Leinster (2002); Leinster (2003); Leinster (2004)]).

2.3.8.4 Categorification

Categorification is the process of finding category-theoretic analogs of set-theoretic concepts by replacing sets with categories, functions with functors,

and equations between functions by natural isomorphisms between functors, which in turn should satisfy certain equations of their own, called ‘coherence laws’. Iterating this process requires a theory of n –categories.

Categorification uses the following analogy between set theory and category theory [Crane and Frenkel (1994); Baez and Dolan (1998)]:

Set Theory	Category Theory
elements	objects
equations between elements	isomorphisms between objects
sets	categories
functions	functors
equations between functions	natural isomorphisms between functors

Just as sets have elements, categories have objects. Just as there are functions between sets, there are functors between categories. Now, the proper analog of an equation between elements is not an equation between objects, but an isomorphism. Similarly, the analog of an equation between functions is a natural isomorphism between functors.

2.3.9 Application: n –Categorical Framework for Higher Gauge Fields

Recall that in the 19th Century, J.C. Maxwell unified Faraday’s electric and magnetic fields. Maxwell’s theory led to Einstein’s special relativity where this unification becomes a spin–off of the unification of space and time in the form of the *Faraday tensor* [Misner *et al.* (1973)]

$$F = E \wedge dt + B,$$

where F is electromagnetic 2–form on space–time, E is electric 1–form on space, and B is magnetic 2–form on space. Gauge theory considers F as secondary object to a connection–potential 1–form A . This makes half of the *Maxwell equations* into tautologies [Baez (2002)], i.e.,

$$F = dA \quad \implies \quad dF = 0 \quad \text{the Bianchi relation,}$$

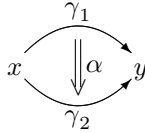
but does not imply the *dual Bianchi relation*, which is a second half of Maxwell’s equations,

$$*d * F = J,$$

where $*$ is the dual *Hodge star operator* and J is current 1-form.

To understand the deeper meaning of the connection-potential 1-form A , we can integrate it along a path γ in space-time, $x \xrightarrow{\gamma} y$. Classically, the integral $\int_{\gamma} A$ represents an *action* for a charged point particle to move along the path γ . Quantum-mechanically, $\exp\left(i \int_{\gamma} A\right)$ represents a *phase* (within the unitary group $U(1)$) by which the particle's wave-function changes as it moves along the path γ , so A is a $U(1)$ -connection.

The only thing that matters here is the *difference* α between two paths γ_1 and γ_2 in the action $\int_{\gamma} A$ [Baez (2002)], which is a two-morphism



To generalize this construction, consider any compact Lie group G . A connection A on a trivial G -bundle is a \mathfrak{g} -valued 1-form. A assigns a *holonomy* $P \exp\left(i \int_{\gamma} A\right) \in G$ along any path $x \xrightarrow{\gamma} y$ and has a *curvature* F given by

$$F = dA + A \wedge A.$$

The curvature F implies the *extended Bianchi relation*

$$dF + A \wedge F = 0,$$

but does not imply the dual Bianchi relation, i.e., *Yang–Mills relation*¹⁵

$$*(d * F + A \wedge * F) = J.$$

¹⁵Recall that the Yang–Mills (YM) Lagrangian density \mathcal{L}_{YM} is a functional of the vector potential fields A^i_μ , where the internal index i ranges over $\{1, \dots, n\}$, where n is the dimension of the gauge group, and μ is a space–time index ($\mu = 0, \dots, 3$). The field tensor derived from these potential fields is (see, e.g., [Pons *et. al.* (2000)])

$$F^i_{\alpha\beta} = A^i_{\beta,\alpha} - A^i_{\alpha,\beta} - C^i_{jk} A^j_\alpha A^k_\beta,$$

where C^i_{jk} are the structure constants of the gauge group Lie. The YM Lagrangian density is consequently given by

$$\mathcal{L}_{\text{YM}} = -\frac{1}{4} \sqrt{|g|} F^i_{\mu\nu} F^j_{\alpha\beta} g^{\mu\alpha} g^{\nu\beta} C_{ij},$$

where C_{ij} is a nonsingular, symmetric group metric and g is the determinant of the space–time metric tensor (in a semi–simple Lie group, C_{ij} is usually taken to be $C^s_{it} C^t_{js}$; in an Abelian Lie group, one usually takes $C_{ij} = \delta_{ij}$).

The derivatives of \mathcal{L}_{YM} with respect to the velocities of the configuration–space variables, \dot{A}^i_α give the tangent–space functions \hat{P}^α_i corresponding to the phase–space conjugate momenta:

$$\hat{P}^\alpha_i = \frac{\partial \mathcal{L}_{\text{YM}}}{\partial \dot{A}^i_\alpha} = \sqrt{|g|} F^j_{\mu\nu} g^{\alpha\mu} g^{0\nu} C_{ij}.$$

The Legendre map \mathcal{FL} is defined by mapping \hat{P}^α_i to P^α_i in the phase–space. Because of the antisymmetry of the field tensor, the primary constraints are

$$0 = \hat{P}^0_i = \frac{\partial \mathcal{L}_{\text{YM}}}{\partial \dot{A}^i_0} = \sqrt{|g|} F^j_{\mu\nu} g^{0\mu} g^{0\nu} C_{ij}.$$

A generator of a projectable gauge transformation thus must be independent of \dot{A}^i_0 .

An infinitesimal YM gauge transformation is defined by an array of gauge fields Λ^i and transforms the potential by

$$\delta_{\text{R}}[\Lambda] A^i_\mu = -\Lambda^i_{,\mu} - C^i_{jk} \Lambda^j A^k_\mu.$$

We denote this transformation by

$$\delta_{\text{R}} A^i_\mu = -(\mathcal{D}_\mu \Lambda)^i,$$

where \mathcal{D}_μ is the *Yang–Mills covariant derivative* (in its action on space–time scalars and YM vectors). Under this transformation, the field transforms as

$$\delta_{\text{R}} F^i_{\mu\nu} = -C^i_{jk} \Lambda^j F^k_{\mu\nu},$$

where we work to first order in Λ^i and use the *Jacobi identity*

$$C^i_{j\ell} C^\ell_{mn} + C^i_{m\ell} C^\ell_{nj} + C^i_{n\ell} C^\ell_{jm} = 0.$$

The YM Lagrangian \mathcal{L}_{YM} is invariant under this transformation provided that the group metric is symmetric,

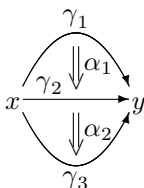
$$C^k_{\ell i} C_{kj} = -C^k_{\ell j} C_{ki} \quad (\text{which is if } C_{ij} = C^s_{it} C^t_{js}).$$

Further generalization is performed with string theory. Just as point particles naturally interact with a 1-form A , strings naturally interact with a 2-form B , such that [Baez (2002)]

$$\text{action} = \int_{\Sigma} B, \quad \text{and} \quad \text{phase} = \exp \left(i \int_{\Sigma} B \right).$$

This 2-form connection B has a 3-form curvature $G = dB$, which satisfies Maxwell-like equations, i.e., implies Bianchi-like relation $dG = 0$, but does not imply the dual, *current relation* $*d*G = J$, with the current 2-form J .

In this way, the *higher Yang-Mills theory* assigns holonomies to paths and also to paths of paths, so that we have a 3-morphism



allowing us to ask not merely whether holonomies along paths are equal, but whether and how they are isomorphic.

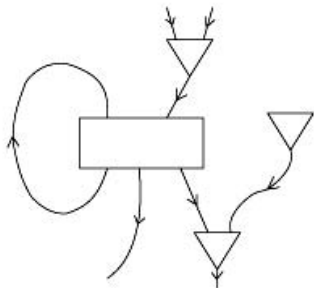
This generalization actually proposes *categorification* of the basic geometrical concepts of manifold, group, Lie group and Lie algebra [Baez (2002)]. Replacing the words *set* and *function* by *smooth manifold* and *smooth map* we get the concept of *smooth category*. Replacing them by *group* and *homomorphism* we get the concept of *2-group*. Replacing them by *Lie group* and *smooth homomorphism* we get the concept of *Lie 2-group*. Replacing them by *Lie algebra* and *Lie algebra homomorphism* we get the concept of *Lie 2-algebra*. Examples of the smooth categories are the following:

- (1) A smooth category with only identity morphisms is a smooth manifold.
- (2) A smooth category with one object and all morphisms invertible is a Lie group.
- (3) Any Lie groupoid gives a smooth category with all morphisms invertible.
- (4) A generalization of a vector bundle (E, M, π) , where E and M are smooth manifolds and projection $\pi : E \rightarrow M$ is a smooth map, gives a vector 2-bundle (E, M, π) where E and M are smooth categories and projection $\pi : E \rightarrow M$ is a smooth functor.

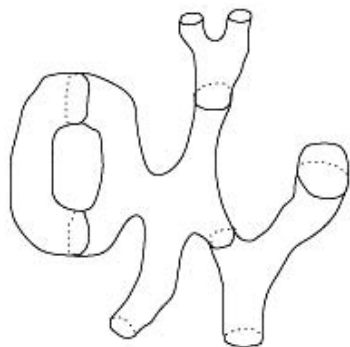
2.3.10 Application: *Natural Geometrical Structures*

Closely related to the higher-dimensional automata are various *natural geometrical structures*, most of which are commonly called tangles.

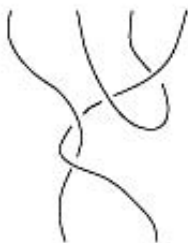
For example, a 2D flow-chart-like complex 1D-structure could be a diagram of the form [Leinster (2002); Leinster (2003)]:



Its 3D-generalization is a surface diagram with the same information-flow:



Moreover, if we allow crossings, as in a *braid*:



then we start getting pictures that look like *knots* which are again related

to higher categorical structures [Leinster (2002); Baez (1997)].

A category \mathcal{C} with only one object is a *monoid* (= semigroup with unit) M . A 2-category \mathcal{C} with only one 0-cell is a *monoidal category* \mathcal{M} . A *braided monoidal category* is a monoidal category equipped with a map called *braiding*

$$A \otimes B \xrightarrow{\beta_{A,B}} B \otimes A,$$

for each pair A, B of objects.

The canonical example of a braided monoidal category is \mathcal{BR} [Leinster (2003)]. This has:

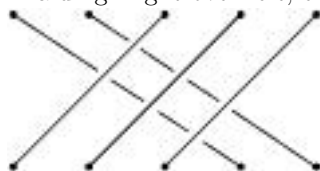
- (1) Objects: natural numbers $0, 1, \dots$;
- (2) Morphisms: braids, e.g.,



(taken up to deformation);

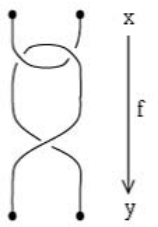
there are no morphisms $m \longrightarrow n$ when $m \neq n$;

- (3) Tensor product: placing side-by-side (which on objects means addition); and
- (4) Braiding: right over left, e.g.,

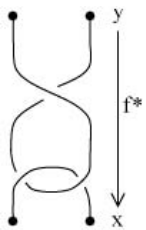


Knots, links and braids are all special cases of *tangles* (see [Reshetikhin and Turaev (1990)]). The mysterious relationships between topology, algebra and physics amount in large part to the existence of interesting functors from various topologically defined categories to **Hilbert**, the category of Hilbert spaces. These topologically defined categories are always **-categories*, and the really interesting functors from them to **Hilbert** are always **-functors*, which preserve the ***-structure. Physically, the *** operation corresponds to *reversing the direction of time*. For example, there is a ***-category whose objects are collections of points and whose morphisms

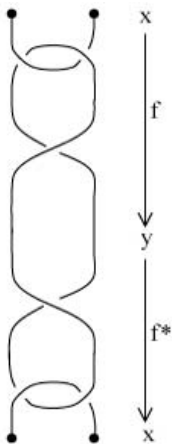
are tangles (see [Baez (1997); Baez and Dolan (1998)]):



We can think of this morphism $f : x \rightarrow y$ as representing the trajectories of a collection of particles and antiparticles, where particles and antiparticles can be created or annihilated in pairs. Reversing the direction of time, we get the ‘dual’ morphism $f^* : y \rightarrow x$:



This morphism is not the inverse of f , since the composite $f \circ f^*$ is a nontrivial tangle:



Indeed, any groupoid becomes a $*$ -category if we set $f^* = f^{-1}$ for every

morphism f .

The above example involves 1D curves in 3D space. More generally, topological quantum field theory studies n D manifolds embedded in $(n+k)$ D space-time, which in the $k \rightarrow \infty$ limit appear as ‘abstract’ n D manifolds. It appears that these are best described using certain ‘ n -categories with duals’, meaning n -categories in which every j -morphism f has a dual f^* .

Therefore, a tangle is a box in 3D space with knotted and linked string embedded within it and a certain number of strands of that string emanating from the surface of the box. There are no open ends of string inside the box. We usually think of some subset of the strands as *inputs* to the tangle and the remaining strands as the *outputs* from the tangle. Usually the inputs are arranged to be drawn vertically and so that they enter tangle from below, while the outputs leave the tangle from above. The tangle itself (within the box) is arranged as nicely as possible with respect to a vertical direction. This means that a definite vertical direction is chosen, and that the tangle intersects planes perpendicular to this direction transversely except for a finite collection of critical points. These basic critical points are local maxima and local minima for the space curves inside the tangle. Two tangles configured with respect to the same box are *ambient isotopic* if there is an isotopy in three space carrying one to the other that fixes the input and output strands of each tangle. We can *compose* two tangles A and B where the number of output strand of A is equal to the number of input strands of B . Composition is accomplished by joining each output strand of A to a corresponding input strand of B [Kauffman and Radford (1995); Kauffman and Radford (1999); Kauffman (1994)].

A *tangle diagram* is a box in the plane, arranged parallel to a chosen vertical direction with a left-right ordered sequence of input strands entering the bottom of the box, and a left-right ordered sequence of output strands emanating from the top of the box. Inside the box is a diagram of the tangle represented with crossings (broken arc indicating the undercrossing line) in the usual way for knot and links. We assume, as above, that the tangle is represented so that it is transverse to lines perpendicular to the vertical except for a finite number of points in the vertical direction along the tangle. It is said that the *tangle is well arranged*, or *Morse* with respect to the vertical direction when these transversality conditions are met. At the critical points we will see a local maximum, a local minimum or a crossing in the diagram. Tangle composition is well-defined (for matching input/output counts) since the input and output strands have an ordering (from left to right for the reader facing the plane on which the tangle di-

agram is drawn). Note that the cardinality of the set of input strands or output strands can be equal to zero. If they are both zero, then the tangle is simply a knot or link diagram arranged well with respect to the vertical direction [Kauffman and Radford (1995); Kauffman and Radford (1999); Kauffman (1994)].

The *Reidemeister moves* are a set of moves on diagrams that combinatorially generate isotopy for knots, links and tangles [Reidemeister (1948)]. If two tangles are equivalent in 3D space, then corresponding diagrams of these tangles can be obtained one from another, by a sequence of Reidemeister moves. Each move is confined to the tangle box and keeps the input and output strands of the tangle diagram fixed.

Two (tangle) diagrams are said to be *regularly isotopic* if one can be obtained from the other by a sequence of Reidemeister moves of type 0,2,3 (move number 1 is not used in regular isotopy).

If A and B are given tangles, we denote the composition of A and B by AB where the diagram of A is placed below the diagram of B and the output strands of A are connected to the input strands of B . If the cardinalities of the sets of input and output strands are zero, then we simply place one tangle below the other to form the product [Kauffman and Radford (1995); Kauffman and Radford (1999); Kauffman (1994)].

Along with tangle composition, as defined in the previous paragraph, we also have an operation of *product* or *juxtaposition* of tangles. To juxtapose two tangles A and B simply place their diagrams side by side with A to the left of B and regard this new diagram as a new tangle whose inputs are the inputs of A followed by the inputs of B , and whose outputs are the outputs of A followed by the outputs of B . We denote the tangle product of A and B by $A \otimes B$.

It remains to describe the equivalence relation on tangles that makes them represent regular isotopy classes of embedded string. Every tangle is a composition of *elementary tangles* where an elementary tangle is one of the following list: a *cup* (a single minimum – zero inputs, two outputs), a *cap* (a single maximum – two inputs, zero outputs), a *crossing* (a single local crossing diagram – two inputs and two outputs).

2.3.11 *Ultimate Conceptual Machines: Weak n –Categories*

As traditionally conceived, an n –category is an algebraic structure having objects or 0–morphisms, 1–morphisms between 0–morphisms,

2-morphisms between 1-morphisms, and so on up to n -morphisms. There should be various ways of composing j -morphisms, and these composition operations should satisfy various laws, such as associativity laws. In the so-called *strict n -categories*, these laws are equations. While well-understood and tractable, strict n -categories are insufficiently general for many applications: what one usually encounters in nature are *weak n -categories*, in which composition operations satisfy the appropriate laws *only up to equivalence*. Here the idea is that n -morphisms are equivalent precisely when they are equal, while for $j < n$ an equivalence between j -morphisms is recursively defined as a $(j + 1)$ -morphism from one to the other that is invertible up to equivalence [Baez and Dolan (1998)].

Now, what makes it difficult to define weak n -categories is that laws formulated as equivalences should satisfy laws of their own – the so-called *coherence laws* – so that one can manipulate them with some of the same facility as equations. Moreover, these coherence laws should also be equivalences satisfying their own coherence laws, again up to equivalence, and so on [Baez and Dolan (1998)].

For example, a weak 1-category is just an ordinary category. In a category, composition of 1-morphisms is associative:

$$(fg)h = f(gh).$$

Weak 2-categories first appeared in the work of Bénabou [Bénabou (1967)], under the name of *bicategories*. In a bicategory, composition of 1-morphisms is associative only up to an invertible 2-morphism, the ‘*associator*’:

$$A_{f,g,h}: (fg)h \rightarrow f(gh).$$

The associator allows one to re-bracket parenthesized composites of arbitrarily many 1-morphisms, but there may be many ways to use it to go from one parenthesization to another. For all these to be equal, the associator must satisfy a coherence law, the *pentagon identity*, which says that the following diagram commutes:

$$\begin{array}{ccc} ((fg)h)i & \xrightarrow{(fg)(hi)} & f(g(hi)) \\ \downarrow & & \uparrow \\ (f(gh))i & \xrightarrow{\quad\quad\quad} & f((gh)i) \end{array}$$

where all the arrows are 2-morphisms built using the associator. Weak 3-categories or *tricategories* were defined by [Gordon *et. al.* (1995)]. In a tricategory, the pentagon identity holds only up to an invertible 3-morphism, which satisfies a further coherence law of its own.

When one explicitly lists the coherence laws this way, the definition of weak n -category tends to grow ever more complicated with increasing n . To get around this, one must carefully study the origin of these coherence laws. So far, most of our insight into coherence laws has been won through homotopy theory, where it is common to impose equations *only up to homotopy*, with these homotopies satisfying coherence laws, again up to homotopy, and so on. For example, the pentagon identity and higher coherence laws for associativity first appeared in Stasheff's work on the structure inherited by a space equipped with a homotopy equivalence to a space with an associative product [Stasheff (1963)]. Subsequent work led to a systematic treatment of coherence laws in homotopy theory through the formalism of *topological operads* [Adams (1978)].

Underlying the connection between homotopy theory and n -category theory is a hypothesis made quite explicit by Grothendieck [Grothendieck (1983)]: to any topological space one should be able to associate an n -category having points as objects, paths between points as 1-morphisms, certain paths of paths as 2-morphisms, and so on, with certain homotopy classes of n -fold paths as n -morphisms. This should be a special sort of weak n -category called a *weak n -groupoid*, in which all j -morphisms ($0 < j \leq n$) are equivalences. Moreover, the process of assigning to each space its *fundamental n -groupoid*, as Grothendieck called it, should set up a complete correspondence between the theory of homotopy n -types (spaces whose homotopy groups vanish above the n th) and the theory of weak n -groupoids. This hypothesis explains why all the coherence laws for weak n -groupoids should be deducible from homotopy theory. It also suggests that weak n -categories will have features not found in homotopy theory, owing to the presence of j -morphisms that are not equivalences [Baez and Dolan (1998)].

Homotopy theory also makes it clear that when setting up a theory of n -categories, there is some choice involved in the shapes of ones j -morphisms – or in the language of topology, *j -cells*. The traditional approach to n -categories is *globular*. This means that for $j > 0$, each j -cell $f: x \rightarrow y$ has two $(j-1)$ -cells called its *source*, $sf = x$, and *target*,

$tf = y$, which for $j > 1$ satisfy

$$s(sf) = s(tf), \quad t(sf) = t(tf).$$

Thus a j -cell can be visualized as a *globe*, a j D ball whose boundary is divided into two $(j - 1)$ D hemispheres corresponding to its source and target. However, in homotopy theory, the simplicial approach is much more popular. In a *simplicial set*, each j -cell f is shaped like a j D simplex, and has $j + 1$ faces, certain $(j - 1)$ -cells d_0f, \dots, d_nf . In addition to these there are $(j + 1)$ -cells $i_0f, \dots, i_{n+1}f$ called *degeneracies*, and the face and degeneracy maps satisfy certain well-known relations [Baez and Dolan (1998)].

This page intentionally left blank

Chapter 3

Applied Manifold Geometry

3.1 Introduction

Albert Einstein once said: “Nature is simple only when analyzed locally. Why? Because, locally any system appears to be linear, and therefore fully predictable and controllable. Geometrical elaboration of this fundamental idea has produced the concept of a *manifold*, a topological space which locally looks like Euclidean \mathbb{R}^n -spaces, but globally can be totally different. In addition, to be able to use calculus on our manifolds, in much the same way as in ordinary \mathbb{R}^n -spaces, the manifolds need to be smooth (i.e., differentiable as many times as required, technically denoted by C^k).

Consider a classical example, comparing a surface of an apple with a Euclidean plane. A small neighborhood of every point on the surface of an apple (excluding its stem) looks like a Euclidean plane (denoted by \mathbb{R}^2), with its *local geodesics* appearing like straight lines. In other words, a smooth surface is *locally topologically equivalent* to the Euclidean plane. This same concept of nonlinear geometry holds in any dimension. If dimension is high, we are dealing with complex systems. Therefore, while continuous-time linear systems live in Euclidean \mathbb{R}^n -spaces, continuous-time complex systems live in n D smooth manifolds, usually denoted by M .

Finally, note that there are two dynamical paradigms of smooth manifolds:

- (i) Einstein’s 4D *space-time manifold*, historically the first one, and
- (ii) n D *configuration manifold*, which is the modern geometrical concept.

As the Einstein space-time manifold is both simpler to comprehend and consequently much more elaborated, we start our geometrical machinery with it, keeping in mind that the same fundamental dynamics holds for all

smooth manifolds, regardless of their dimension.

Throughout the book we will try to follow the Hilbertian pedagogical method of development: (i) intuitively introduce a new geometrical concept; (ii) rigorously define it; (iii) apply it to solve a real-world problem.

3.1.1 Intuition Behind Einstein's Geometrodynamics

Briefly, Einstein–Wheeler geometrodynamics can be summarized as [Wheeler (1961); Wheeler (1962)]:

- (1) Gravity is not a Newtonian force, but an aspect of the geometry of space–time.
- (2) Space is not an absolute invariant entity, but is influenced by the distribution of mass and energy in the Universe. The fundamental Geometrodynamics Principle states:
Space tells matter how to move, while matter tells space how to curve.
- (3) Large masses introduce a strong curvature in space–time. Light and matter are forced to move according to this metric. Since all the matter is in motion, the geometry of space is constantly changing.

The celebrated *Einstein equation* relates the curvature of space–time to the mass/energy density. It reads (in the so-called ‘normal’ units: $c = 8\pi G = 1$):

$$\mathbf{G} = \mathbf{T}, \quad \text{or, in components,} \quad G_{\alpha\beta} = T_{\alpha\beta}, \quad (3.1)$$

where $\mathbf{G} = G_{\alpha\beta}$ is the Einstein curvature tensor, representing space–time geometry, while $\mathbf{T} = T_{\alpha\beta}$ is the stress–energy–momentum tensor, the ‘mystical’ SEM-tensor, representing matter; the 4D indices $\alpha, \beta = (0, 1, 2, 3)$ label respectively the four space–time directions: (t, x, y, z) .

To grasp the intuition behind the Einstein equation (3.1), we need to consider a ball filled with test particles that are all initially at rest relative to each other. Let $V = V(t)$ be the volume of the ball after a proper time t has elapsed, as measured by the particle at the center of the ball. Then the Einstein equation says:

$$\left. \frac{\ddot{V}}{V} \right|_{t=0} = -\frac{1}{2} \begin{pmatrix} \text{flow of } t\text{-momentum in } t\text{-direction} + \\ \text{flow of } x\text{-momentum in } x\text{-direction} + \\ \text{flow of } y\text{-momentum in } y\text{-direction} + \\ \text{flow of } z\text{-momentum in } z\text{-direction} \end{pmatrix},$$

where these flows are measured at the center of the ball at time $t = 0$, using local inertial coordinates. These flows are the diagonal components of the SEM-tensor \mathbf{T} . Its components $T_{\alpha\beta}$ tell us how much momentum in the α -direction is flowing in the β -direction through a given point of space-time. The flow of t -momentum in the t -direction is just the energy density, $T_{00} = \rho$. The flow of x -momentum in the x -direction is the ‘pressure in the x -direction’, $T_{11} = P_1 \equiv P_x$, and similarly for y and z .

In any event, we may summarize the Einstein equation (3.1) as

$$\left. \frac{\ddot{V}}{V} \right|_{t=0} = -\frac{1}{2}(\rho + P_x + P_y + P_z) \equiv -\frac{1}{2}(T_{00} + T_{11} + T_{22} + T_{33}). \quad (3.2)$$

This new equation tells us that positive energy density and positive pressure curve space-time in a way that makes a freely falling ball of point particles tend to shrink. Since $E = mc^2$ and we are working in normal units, ordinary mass density counts as a form of energy density. Thus a massive object will make a swarm of freely falling particles at rest around it start to shrink. In short, (3.2) tells us that *gravity attracts* (see e.g., [Misner *et al.* (1973); Baez (2001)]).

To see why equation (3.2) is equivalent to the Einstein equation (3.1), we need to understand the *Riemann curvature tensor* and the *geodesic deviation equation*. Namely, when space-time is curved, the result of *parallel transport* depends on the path taken. To quantify this notion, pick two vectors u and v at a point p in space-time. In the limit where $\epsilon \rightarrow 0$, we can approximately speak of a ‘parallelogram’ with sides ϵu and ϵv . Take another vector w at p and parallel transport it first along ϵv and then along ϵu to the opposite corner of this parallelogram. The result is some vector w_1 . Alternatively, parallel transport w first along ϵu and then along ϵv . The result is a slightly different vector, w_2 . The limit

$$\lim_{\epsilon \rightarrow 0} \frac{w_2 - w_1}{\epsilon^2} = R(u, v)w \quad (3.3)$$

is well-defined, and it measures the curvature of space-time at the point p . In local coordinates, we can write it as

$$R(u, v)w = R_{\beta\gamma\delta}^{\alpha} u^{\beta} v^{\gamma} w^{\delta}.$$

The quantity $R_{\beta\gamma\delta}^{\alpha}$ is called the *Riemann curvature tensor*. We can use this tensor to calculate the relative acceleration of nearby particles in free fall if they are initially at rest relative to one another. Consider two freely falling particles at nearby points p and q . Let v be the velocity of the particle at

p , and let ϵu be the vector from p to q . Since the two particles start out at rest relative to one other, the velocity of the particle at q is obtained by parallel transporting v along ϵu .

Now let us wait a short while. Both particles trace out geodesics as time passes, and at time ϵ they will be at new points, say p' and q' . The point p' is displaced from p by an amount ϵv , so we get a little parallelogram, exactly as in the definition of the Riemann curvature:

Next let us calculate the new relative velocity of the two particles. To compare vectors we must carry one to another using parallel transport. Let v_1 be the vector we get by taking the velocity vector of the particle at p' and parallel transporting it to q' along the top edge of our parallelogram. Let v_2 be the velocity of the particle at q' . The difference $v_2 - v_1$ is the new relative velocity. It follows that over this passage of time, the average relative acceleration of the two particles is $a = (v_2 - v_1)/\epsilon$. By equation (3.3),

$$\lim_{\epsilon \rightarrow 0} \frac{v_2 - v_1}{\epsilon^2} = R(u, v)v, \quad \text{therefore} \quad \lim_{\epsilon \rightarrow 0} \frac{a}{\epsilon} = R(u, v)v.$$

This is the simplified form of the *geodesic deviation equation*. From the definition of the Riemann curvature it is easy to see that $R(u, v)w = -R(v, u)w$, so we can also write this equation as

$$\lim_{\epsilon \rightarrow 0} \frac{a^\alpha}{\epsilon} = -R_{\beta\gamma\delta}^\alpha v^\beta u^\gamma v^\delta. \quad (3.4)$$

Using geodesic deviation equation (3.4) we can work out the second time derivative of the volume $V(t)$ of a small ball of test particles that start out at rest relative to each other. For this we must let u range over an orthonormal basis of tangent vectors, and sum the ‘outwards’ component of acceleration for each one of these. By equation (3.4) this gives

$$\lim_{V \rightarrow 0} \frac{\ddot{V}}{V} \Big|_{t=0} = -R_{\beta\alpha\delta}^\alpha v^\beta v^\delta.$$

In terms of the so-called *Ricci tensor*, which is a contracted Riemann tensor,

$$R_{\beta\delta} = R_{\beta\alpha\delta}^\alpha,$$

we may write the above expression as

$$\lim_{V \rightarrow 0} \frac{\ddot{V}}{V} \Big|_{t=0} = -R_{\beta\delta} v^\beta v^\delta.$$

In local inertial coordinates, where the ball starts out at rest, we have $v = (1, 0, 0, 0)$, so

$$\lim_{V \rightarrow 0} \frac{\ddot{V}}{V} \Big|_{t=0} = -R_{00}. \quad (3.5)$$

In short, the Ricci tensor says how our ball of freely falling test particles starts changing in volume. The Ricci tensor only captures some of the information in the *Riemann curvature tensor*. The rest is captured by the so-called the *Weyl tensor* (see e.g., [Penrose (1989); Penrose (1994); Penrose (1997)]), which says how any such ball starts changing in shape. The Weyl tensor describes tidal forces, gravitational waves and the like.

Now, the Einstein equation in its usual form says

$$G_{\alpha\beta} = T_{\alpha\beta}.$$

Here the right side is the stress-energy tensor, while the left side, the ‘Einstein tensor’, is just an abbreviation for a quantity constructed from the Ricci tensor:

$$G_{\alpha\beta} = R_{\alpha\beta} - \frac{1}{2}g_{\alpha\beta}R^\gamma_\gamma.$$

Thus the Einstein equation really says

$$R_{\alpha\beta} - \frac{1}{2}g_{\alpha\beta}R^\gamma_\gamma = T_{\alpha\beta}. \quad (3.6)$$

This implies

$$R^\alpha_\alpha - \frac{1}{2}g^\alpha_\alpha R^\gamma_\gamma = T^\alpha_\alpha,$$

but $g^\alpha_\alpha = 4$, so

$$-R^\alpha_\alpha = T^\alpha_\alpha.$$

Substituting this into equation (3.6), we get

$$R_{\alpha\beta} = T_{\alpha\beta} - \frac{1}{2}g_{\alpha\beta}T^\gamma_\gamma. \quad (3.7)$$

This is an equivalent version of the Einstein equation, but with the roles of R and T switched [Baez (2001)]. This is a formula for the Ricci tensor, which has a simple geometrical meaning.

Equation (3.7) will be true if any one component holds in all local inertial coordinate systems. This is a bit like the observation that all of Maxwell’s equations are contained in Gauss’s law and $\nabla \cdot B = 0$.

Clearly, this is only true if we know how the fields transform under change of coordinates. Here we assume that the transformation laws are known. Given this, the Einstein equation (3.1) is equivalent to the fact that

$$R_{00} = T_{00} - \frac{1}{2}g_{00}T_{\gamma}^{\gamma} \quad (3.8)$$

in every local inertial coordinate system about every point. In such coordinates we have

$$g = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (3.9)$$

so $g_{00} = -1$, as well as

$$T_{\gamma}^{\gamma} = -T_{00} + T_{11} + T_{22} + T_{33}.$$

Equation (3.8) thus says that

$$R_{00} = \frac{1}{2}(T_{00} + T_{11} + T_{22} + T_{33}).$$

By equation (3.5), this is equivalent to the required

$$\lim_{V \rightarrow 0} \frac{\ddot{V}}{V} \Big|_{t=0} = -\frac{1}{2}(T_{00} + T_{11} + T_{22} + T_{33}).$$

3.1.2 Einstein's Geometrodynamics in Brief

As a final introductory motivation, we give an ‘express-flight bird-view’ on *derivation of the Einstein equation* from the *Hilbert action principle*, starting from the Einstein space-time manifold M . For all technical details, see [Misner *et al.* (1973)], which is still, after 33 years, the core textbook

on the subject.

$$\begin{aligned}
M & \quad \dots \text{ the } \mathbf{space-time manifold } M \\
g_{ij} &= g_{ij}(x^i) \in T_x M \quad \dots \text{ metric tensor on } M \\
g^{ij} &= (g_{ij})^{-1} \quad \dots \text{ inverse metric tensor on } M \\
\Gamma_{ijk} &= \frac{1}{2}(\partial_{x^k} g_{ij} + \partial_{x^j} g_{ki} - \partial_{x^i} g_{jk}) \quad \dots \text{ 1-order Christoffel symbols} \\
\Gamma_{ij}^k &= g^{kl} \Gamma_{ijl} \quad \dots \text{ 2-order Christoffel symbols (Levi-Civita connection)} \\
R_{ijk}^l &= \partial_{x^j} \Gamma_{ik}^l - \partial_{x^k} \Gamma_{ij}^l + \Gamma_{rj}^l \Gamma_{ik}^r - \Gamma_{rk}^l \Gamma_{ij}^r \quad \dots \text{ Riemann curvature tensor} \\
R_{ij} &= R_{ijl}^l \quad \dots \text{ Ricci tensor is the trace of Riemann} \\
R &= g^{ij} R_{ij} \quad \dots \text{ scalar curvature is the trace of Ricci} \\
G_{ij} &= R_{ij} - \frac{1}{2} R g_{ij} \quad \dots \text{ Einstein tensor is the trace-reversed Ricci} \\
T_{ij} &= -2 \frac{\delta L_{Hilb}}{\delta g^{ij}} + g_{ij} L_{Hilb} \quad \dots \text{ stress-energy-momentum (SEM) tensor} \\
L_{Hilb} &= \frac{1}{16\pi} g^{ij} R_{ij} (-g)^{1/2} \quad \dots \text{ is derived from the Hilbert Lagrangian} \\
\delta S &= \delta \int L_{Hilb} (-g)^{1/2} d^4 x = 0 \quad \dots \text{ the Hilbert action principle gives} \\
G_{ij} &= 8\pi T_{ij} \quad \dots \text{ the } \mathbf{Einstein equation}.
\end{aligned}$$

We will continue Einstein's geometrodynamics in section 6.4 below.

3.2 Intuition Behind the Manifold Concept

As we have already got the initial feeling, in the heart of applied differential geometry is the concept of a *manifold*. As a warm-up, to get some *dynamical intuition* behind this concept, let us consider a simple 3DOF mechanical system determined by three generalized coordinates, $q^i = \{q^1, q^2, q^3\}$. There is a unique way to represent this system as a 3D manifold, such that to each point of the manifold there corresponds a definite configuration of the mechanical system with coordinates q^i ; therefore, we have a geometrical representation of the configurations of our mechanical system, called the *configuration manifold*. If the mechanical system moves in any way, its coordinates are given as the functions of the time. Thus, the motion is given by equations of the form: $q^i = q^i(t)$. As t varies (i.e., $t \in \mathbb{R}$), we observe that the system's *representative point* in the configuration manifold describes a *curve* and $\dot{q}^i = \dot{q}^i(t)$ are the equations of this curve.

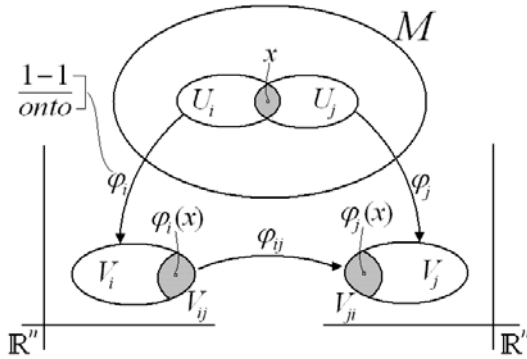


Fig. 3.1 An intuitive geometrical picture behind the manifold concept (see text).

On the other hand, to get some *geometrical intuition* behind the concept of a manifold, consider a set M (see Figure 3.1) which is a *candidate* for a manifold. Any point $x \in M^1$ has its *Euclidean chart*, given by a 1–1 and *onto* map $\varphi_i : M \rightarrow \mathbb{R}^n$, with its *Euclidean image* $V_i = \varphi_i(U_i)$. More precisely, a chart φ_i is defined by

$$\varphi_i : M \supset U_i \ni x \mapsto \varphi_i(x) \in V_i \subset \mathbb{R}^n,$$

where $U_i \subset M$ and $V_i \subset \mathbb{R}^n$ are open sets (see [Boothby (1986); Arnold (1978); De Rham (1984)]).

Clearly, any point $x \in M$ can have several different charts (see Figure 3.1). Consider a case of two charts, $\varphi_i, \varphi_j : M \rightarrow \mathbb{R}^n$, having in their images two open sets, $V_{ij} = \varphi_i(U_i \cap U_j)$ and $V_{ji} = \varphi_j(U_i \cap U_j)$. Then we have *transition functions* φ_{ij} between them,

$$\varphi_{ij} = \varphi_j \circ \varphi_i^{-1} : V_{ij} \rightarrow V_{ji}, \quad \text{locally given by} \quad \varphi_{ij}(x) = \varphi_j(\varphi_i^{-1}(x)).$$

If transition functions φ_{ij} exist, then we say that two charts, φ_i and φ_j are *compatible*. Transition functions represent a general (nonlinear) *transformations of coordinates*, which are the core of classical *tensor calculus*.

A set of compatible charts $\varphi_i : M \rightarrow \mathbb{R}^n$, such that each point $x \in M$ has its Euclidean image in at least one chart, is called an *atlas*. Two atlases are *equivalent* iff all their charts are compatible (i.e., transition functions

¹Note that sometimes we will denote the point in a manifold M by m , and sometimes by x (thus implicitly assuming the existence of coordinates $x = (x^i)$).

exist between them), so their union is also an atlas. A *manifold structure* is a class of equivalent atlases.

Finally, as charts $\varphi_i : M \rightarrow \mathbb{R}^n$ were supposed to be 1-1 and onto maps, they can be either *homeomorphisms*, in which case we have a *topological* (C^0) manifold, or *diffeomorphisms*, in which case we have a *smooth* (C^k) manifold.

Slightly more precisely, a topological (respectively smooth) manifold is a separable space M which is locally homeomorphic (resp. diffeomorphic) to Euclidean space \mathbb{R}^n , having the following properties (reflected in Figure 3.1):

- (1) M is a *Hausdorff space*: For every pair of points $x_1, x_2 \in M$, there are disjoint open subsets $U_1, U_2 \subset M$ such that $x_1 \in U_1$ and $x_2 \in U_2$.
- (2) M is *second-countable space*: There exists a countable basis for the topology of M .
- (3) M is *locally Euclidean of dimension n* : Every point of M has a neighborhood that is homeomorphic (resp. diffeomorphic) to an open subset of \mathbb{R}^n .

This implies that for any point $x \in M$ there is a homeomorphism (resp. diffeomorphism) $\varphi : U \rightarrow \varphi(U) \subseteq \mathbb{R}^n$, where U is an open neighborhood of x in M and $\varphi(U)$ is an open subset in \mathbb{R}^n . The pair (U, φ) is called a *coordinate chart* at a point $x \in M$, etc.

3.3 Definition of a Differentiable Manifold

Given a chart (U, φ) , we call the set U a *coordinate domain*, or a coordinate neighborhood of each of its points. If in addition $\varphi(U)$ is an open ball in \mathbb{R}^n , then U is called a *coordinate ball*. The map φ is called a (*local*) *coordinate map*, and the component functions (x^1, \dots, x^n) of φ , defined by $\varphi(m) = (x^1(m), \dots, x^n(m))$, are called *local coordinates* on U .

Two charts (U_1, φ_1) and (U_2, φ_2) such that $U_1 \cap U_2 \neq \emptyset$ are called *compatible* if $\varphi_1(U_1 \cap U_2)$ and $\varphi_2(U_2 \cap U_1)$ are open subsets of \mathbb{R}^n . A family $(U_\alpha, \varphi_\alpha)_{\alpha \in A}$ of compatible charts on M such that the U_α form a *covering* of M is called an *atlas*. The maps $\varphi_{\alpha\beta} = \varphi_\beta \circ \varphi_\alpha^{-1} : \varphi_\alpha(U_{\alpha\beta}) \rightarrow \varphi_\beta(U_{\alpha\beta})$ are called the *transition maps*, for the atlas $(U_\alpha, \varphi_\alpha)_{\alpha \in A}$, where $U_{\alpha\beta} = U_\alpha \cap U_\beta$,

so that we have a commutative triangle:

$$\begin{array}{ccc}
 & U_{\alpha\beta} \subseteq M & \\
 \varphi_\alpha \swarrow & & \searrow \varphi_\beta \\
 \varphi_\alpha(U_{\alpha\beta}) & \xrightarrow{\varphi_{\alpha\beta}} & \varphi_\beta(U_{\alpha\beta})
 \end{array}$$

An atlas $(U_\alpha, \varphi_\alpha)_{\alpha \in A}$ for a manifold M is said to be a C^k -atlas, if all transition maps $\varphi_{\alpha\beta} : \varphi_\alpha(U_{\alpha\beta}) \rightarrow \varphi_\beta(U_{\alpha\beta})$ are of class C^k . Two C^k atlases are called C^k -equivalent, if their union is again a C^k -atlas for M . An equivalence class of C^k -atlases is called a C^k -structure on M . In other words, a smooth structure on M is a *maximal* smooth atlas on M , i.e., such an atlas that is not contained in any strictly larger smooth atlas. By a C^k -manifold M , we mean a topological manifold together with a C^k -structure and a chart on M will be a chart belonging to some atlas of the C^k -structure. Smooth manifold means C^∞ -manifold, and the word ‘smooth’ is used synonymously with C^∞ [De Rham (1984)].

Sometimes the terms ‘local coordinate system’ or ‘parametrization’ are used instead of charts. That M is not defined with any particular atlas, but with an equivalence class of atlases, is a mathematical formulation of the *general covariance* principle. Every suitable coordinate system is equally good. A Euclidean chart may well suffice for an open subset of \mathbb{R}^n , but this coordinate system is not to be preferred to the others, which may require many charts (as with polar coordinates), but are more convenient in other respects.

For example, the atlas of an n -sphere S^n has two charts. If $N = (1, 0, \dots, 0)$ and $S = (-1, \dots, 0, 0)$ are the north and south poles of S^n respectively, then the two charts are given by the stereographic projections from N and S :

$$\begin{aligned}
 \varphi_1 : S^n \setminus \{N\} &\rightarrow \mathbb{R}^n, \varphi_1(x^1, \dots, x^{n+1}) = (x^2/(1-x^1), \dots, x^{n+1}/(1-x^1)), \text{ and} \\
 \varphi_2 : S^n \setminus \{S\} &\rightarrow \mathbb{R}^n, \varphi_2(x^1, \dots, x^{n+1}) = (x^2/(1+x^1), \dots, x^{n+1}/(1+x^1)),
 \end{aligned}$$

while the overlap map $\varphi_2 \circ \varphi_1^{-1} : \mathbb{R}^n \setminus \{0\} \rightarrow \mathbb{R}^n \setminus \{0\}$ is given by the diffeomorphism $(\varphi_2 \circ \varphi_1^{-1})(z) = z/||z||^2$, for z in $\mathbb{R}^n \setminus \{0\}$, from $\mathbb{R}^n \setminus \{0\}$ to itself.

Various *additional structures* can be imposed on \mathbb{R}^n , and the corresponding manifold M will inherit them through its covering by charts. For example, if a covering by charts takes their values in a *Banach space* E ,

then E is called the *model space* and M is referred to as a C^k -*Banach manifold* modelled on E . Similarly, if a covering by charts takes their values in a *Hilbert space* \mathcal{H} , then \mathcal{H} is called the *model space* and M is referred to as a C^k -*Hilbert manifold* modelled on \mathcal{H} . If not otherwise specified, we will consider M to be an Euclidean manifold, with its covering by charts taking their values in \mathbb{R}^n .

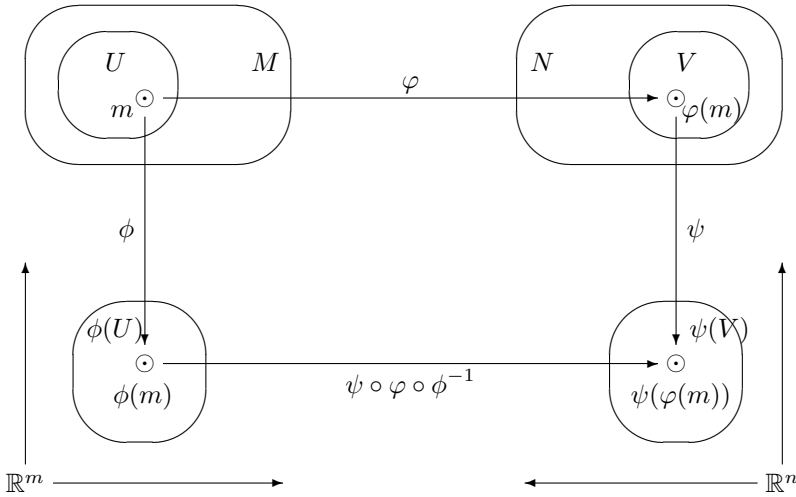
For a Hausdorff C^k -manifold the following properties are equivalent [Kolar *et al.* (1993)]: (i) it is paracompact; (ii) it is metrizable; (iii) it admits a Riemannian metric;² (iv) each connected component is separable.

3.4 Smooth Maps Between Smooth Manifolds

A map $\varphi : M \rightarrow N$ between two manifolds M and N , with $M \ni m \mapsto \varphi(m) \in N$, is called a *smooth map*, or C^k -map, if we have the following charting:

²Recall the corresponding properties of a *Euclidean metric* d . For any three points $x, y, z \in \mathbb{R}^n$, the following axioms are valid:

$$\begin{array}{ll} M_1 : d(x, y) > 0, & \text{for } x \neq y; \quad \text{and} \quad d(x, y) = 0, \quad \text{for } x = y; \\ M_2 : d(x, y) = d(y, x); & \\ M_3 : d(x, y) \leq d(x, z) + d(z, y). & \end{array}$$



This means that for each $m \in M$ and each chart (V, ψ) on N with $\varphi(m) \in V$ there is a chart (U, ϕ) on M with $m \in U$, $\varphi(U) \subseteq V$, and $\Phi = \psi \circ \varphi \circ \phi^{-1}$ is C^k , that is, the following diagram commutes:

$$\begin{array}{ccc}
 M \supseteq U & \xrightarrow{\varphi} & V \subseteq N \\
 \phi \downarrow & & \downarrow \psi \\
 \phi(U) & \xrightarrow{\Phi} & \psi(V)
 \end{array}$$

Let M and N be smooth manifolds and let $\varphi : M \rightarrow N$ be a smooth map. The map φ is called a *covering*, or equivalently, M is said to *cover* N , if φ is surjective and each point $n \in N$ admits an open neighborhood V such that $\varphi^{-1}(V)$ is a union of disjoint open sets, each diffeomorphic via φ to V .

A C^k -map $\varphi : M \rightarrow N$ is called a C^k -*diffeomorphism* if φ is a bijection, $\varphi^{-1} : N \rightarrow M$ exists and is also C^k . Two manifolds are called diffeomorphic if there exists a diffeomorphism between them. All smooth manifolds and smooth maps between them form the category \mathcal{M} .

3.4.1 Intuition Behind Topological Invariants of Manifolds

Now, restricting to the topology of n D compact (i.e., closed and bounded) and connected manifolds, the only cases in which we have a complete understanding of topology are $n = 0, 1, 2$. The only compact and connected

0D manifold is a point. A 1D compact and connected manifold can either be a line element or a circle, and it is intuitively clear (and can easily be proven) that these two spaces are topologically different. In 2D, there is already an infinite number of different topologies: a 2D compact and connected surface can have an arbitrary number of handles and boundaries, and can either be orientable or non-orientable (see figure 3.2). Again, it is intuitively quite clear that two surfaces are not homeomorphic if they differ in one of these respects. On the other hand, it can be proven that any two surfaces for which these data are the same can be continuously mapped to one another, and hence this gives a complete classification of the possible topologies of such surfaces.

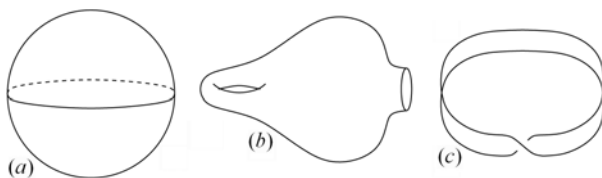


Fig. 3.2 Three examples of 2D manifolds: (a) The sphere S^2 is an orientable manifold without handles or boundaries. (b) An orientable manifold with one boundary and one handle. (c) The Möbius strip is an unorientable manifold with one boundary and no handles.

A quantity such as the number of boundaries of a surface is called a *topological invariant*. A topological invariant is a number, or more generally any type of structure, which one can associate to a topological space, and which does not change under continuous mappings. Topological invariants can be used to distinguish between topological spaces: if two surfaces have a different number of boundaries, they can certainly not be topologically equivalent. On the other hand, the knowledge of a topological invariant is in general not enough to decide whether two spaces are homeomorphic: a torus and a sphere have the same number of boundaries (zero), but are clearly not homeomorphic. Only when one has some *complete set* of topological invariants, such as the number of handles and boundaries in the 2D case, is it possible to determine whether or not two topological spaces are homeomorphic. In more than 2D, many topological invariants are known, but for no dimension larger than two has a complete set of topological invariants been found. In 3D, it is generally believed that a finite number of countable invariants would suffice for compact manifolds, but this is not rig-

ously proven, and in particular there is at present no generally accepted construction of a complete set. A very interesting and intimately related problem is the famous *Poincaré conjecture*, stating that if a 3D manifold has a certain set of topological invariants called its ‘homotopy groups’ equal to those of the 3-sphere S^3 , it is actually homeomorphic to the three-sphere. In four or more dimensions, a complete set of topological invariants would consist of an uncountably infinite number of invariants! A general classification of topologies is therefore very hard to get, but even without such a general classification, each new invariant that can be constructed gives us a lot of interesting new information. For this reason, the construction of topological invariants of manifolds is one of the most important issues in topology.

3.5 (Co)Tangent Bundles of Smooth Manifolds

3.5.1 Tangent Bundle and Lagrangian Dynamics

3.5.1.1 Intuition Behind a Tangent Bundle

In mechanics, to each n D *configuration manifold* M there is associated its $2n$ D *velocity phase-space manifold*, denoted by TM and called the tangent bundle of M (see Figure 3.3). The original smooth manifold M is called the *base* of TM . There is an onto map $\pi : TM \rightarrow M$, called the *projection*. Above each point $x \in M$ there is a *tangent space* $T_x M = \pi^{-1}(x)$ to M at x , which is called a *fibre*. The fibre $T_x M \subset TM$ is the subset of TM , such that the total tangent bundle, $TM = \bigsqcup_{m \in M} T_x M$, is a *disjoint union* of tangent spaces $T_x M$ to M for all points $x \in M$. From dynamical perspective, the most important quantity in the tangent bundle concept is the smooth map $v : M \rightarrow TM$, which is an inverse to the projection π , i.e. $\pi \circ v = \text{Id}_M$, $\pi(v(x)) = x$. It is called the *velocity vector-field*. Its graph $(x, v(x))$ represents the *cross-section* of the tangent bundle TM . This explains the dynamical term velocity phase-space, given to the tangent bundle TM of the manifold M .

3.5.1.2 Definition of a Tangent Bundle

Recall that if $[a, b]$ is a closed interval, a C^0 -map $\gamma : [a, b] \rightarrow M$ is said to be *differentiable* at the endpoint a if there is a chart (U, ϕ) at $\gamma(a)$ such

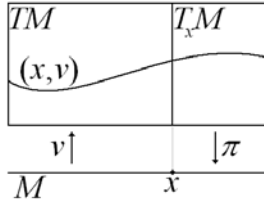


Fig. 3.3 A sketch of a tangent bundle TM of a smooth manifold M (see text for explanation).

that the following limit exists and is finite [Abraham *et al.* (1988)]:

$$\frac{d}{dt}(\phi \circ \gamma)(a) \equiv (\phi \circ \gamma)'(a) = \lim_{t \rightarrow a} \frac{(\phi \circ \gamma)(t) - (\phi \circ \gamma)(a)}{t - a}. \quad (3.10)$$

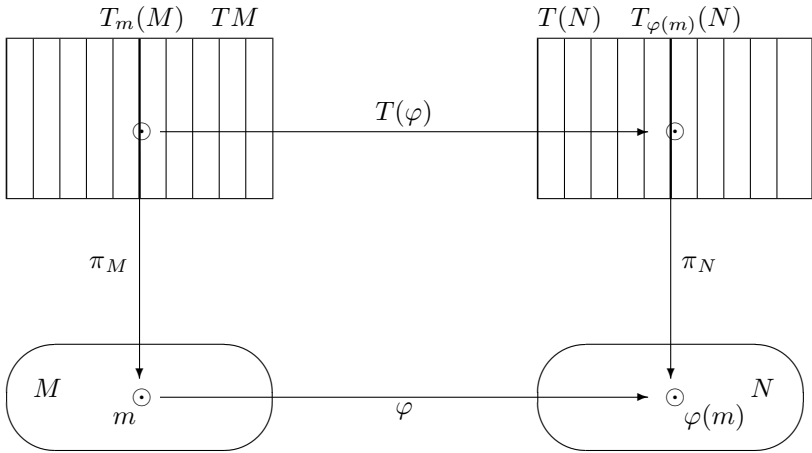
Generalizing (3.10), we get the notion of the *curve on a manifold*. For a smooth manifold M and a point $m \in M$ a curve at m is a C^0 -map $\gamma : I \rightarrow M$ from an interval $I \subset \mathbb{R}$ into M with $0 \in I$ and $\gamma(0) = m$.

Two curves γ_1 and γ_2 passing through a point $m \in U$ are *tangent at m* with respect to the chart (U, ϕ) if $(\phi \circ \gamma_1)'(0) = (\phi \circ \gamma_2)'(0)$. Thus, two curves are tangent if they have identical tangent vectors (same direction and speed) in a local chart on a manifold.

For a smooth manifold M and a point $m \in M$, the *tangent space* $T_m M$ to M at m is the *set of equivalence classes* of curves at m :

$$T_m M = \{[\gamma]_m : \gamma \text{ is a curve at a point } m \in M\}.$$

A C^k -map $\varphi : M \ni m \mapsto \varphi(m) \in N$ between two manifolds M and N induces a linear map $T_m \varphi : T_m M \rightarrow T_{\varphi(m)} N$ for each point $m \in M$, called a *tangent map*, if we have:



i.e., the following diagram commutes:

$$\begin{array}{ccc}
 T_m M & \xrightarrow{T_m \varphi} & T_{\varphi(m)} N \\
 \pi_M \downarrow & & \downarrow \pi_N \\
 M \ni m & \xrightarrow{\varphi} & \varphi(m) \in N
 \end{array}$$

with the *natural projection* $\pi_M : TM \rightarrow M$, given by $\pi_M(T_m M) = m$, that takes a tangent vector v to the point $m \in M$ at which the vector v is attached i.e., $v \in T_m M$.

For an n D smooth manifold M , its n D *tangent bundle* TM is the disjoint union of all its tangent spaces $T_m M$ at all points $m \in M$, $TM = \bigsqcup_{m \in M} T_m M$.

To define the smooth structure on TM , we need to specify how to construct local coordinates on TM . To do this, let $(x^1(m), \dots, x^n(m))$ be local coordinates of a point m on M and let $(v^1(m), \dots, v^n(m))$ be components of a tangent vector in this coordinate system. Then the $2n$ numbers $(x^1(m), \dots, x^n(m), v^1(m), \dots, v^n(m))$ give a *local coordinate system* on TM .

$TM = \bigsqcup_{m \in M} T_m M$ defines a family of vector spaces parameterized by M .

The inverse image $\pi_M^{-1}(m)$ of a point $m \in M$ under the natural projection π_M is the tangent space $T_m M$. This space is called the *fibre* of the tangent bundle over the point $m \in M$ [Steenrod (1951)].

A C^k -map $\varphi : M \rightarrow N$ between two manifolds M and N induces a linear *tangent map* $T\varphi : TM \rightarrow TN$ between their tangent bundles, i.e.,

the following diagram commutes:

$$\begin{array}{ccc}
 TM & \xrightarrow{T\varphi} & TN \\
 \pi_M \downarrow & & \downarrow \pi_N \\
 M & \xrightarrow{\varphi} & N
 \end{array}$$

All tangent bundles and their tangent maps form the category \mathcal{TB} . The category \mathcal{TB} is the natural framework for *Lagrangian dynamics*.

Now, we can formulate the *global version of the chain rule*. If $\varphi : M \rightarrow N$ and $\psi : N \rightarrow P$ are two smooth maps, then we have $T(\psi \circ \varphi) = T\psi \circ T\varphi$ (see [Kolar *et al.* (1993)]). In other words, we have a functor $T : \mathcal{M} \Rightarrow \mathcal{TB}$ from the category \mathcal{M} of smooth manifolds to the category \mathcal{TB} of their tangent bundles:

$$\begin{array}{ccc}
 M & & TM \\
 \varphi \swarrow & & \swarrow T\varphi \\
 N & \xrightarrow{\psi} & P \\
 & \searrow (\psi \circ \varphi) & \searrow T(\psi \circ \varphi) \\
 & & TP
 \end{array}
 \quad \xRightarrow{T} \quad
 \begin{array}{ccc}
 TM & & TP \\
 T\varphi \swarrow & & \swarrow T(\psi \circ \varphi) \\
 TN & \xrightarrow{T\psi} & TP
 \end{array}$$

3.5.2 Cotangent Bundle and Hamiltonian Dynamics

3.5.2.1 Definition of a Cotangent Bundle

A *dual* notion to the tangent space $T_m M$ to a smooth manifold M at a point m is its *cotangent space* $T_m^* M$ at the same point m . Similarly to the tangent bundle, for a smooth manifold M of dimension n , its *cotangent bundle* $T^* M$ is the disjoint union of all its cotangent spaces $T_m^* M$ at all points $m \in M$, i.e., $T^* M = \bigsqcup_{m \in M} T_m^* M$. Therefore, the cotangent bundle

of an n -manifold M is the vector bundle $T^* M = (TM)^*$, the (real) dual of the tangent bundle TM .

If M is an n -manifold, then $T^* M$ is a $2n$ -manifold. To define the smooth structure on $T^* M$, we need to specify how to construct local coordinates on $T^* M$. To do this, let $(x^1(m), \dots, x^n(m))$ be local coordinates of a point m on M and let $(p_1(m), \dots, p_n(m))$ be components of a covector in this coordinate system. Then the $2n$ numbers $(x^1(m), \dots, x^n(m), p_1(m), \dots, p_n(m))$ give a local coordinate system on $T^* M$. This is the basic idea one uses to prove that indeed $T^* M$ is a

$2n$ -manifold.

$T^*M = \bigsqcup_{m \in M} T_m^*M$ defines a family of vector spaces parameterized by M , with the *conatural projection*, $\pi_M^* : T^*M \rightarrow M$, given by $\pi_M^*(T_m^*M) = m$, that takes a covector p to the point $m \in M$ at which the covector p is attached i.e., $p \in T_m^*M$. The inverse image $\pi_M^{-1}(m)$ of a point $m \in M$ under the conatural projection π_M^* is the cotangent space T_m^*M . This space is called the *fibre* of the cotangent bundle over the point $m \in M$.

In a similar way, a C^k -map $\varphi : M \rightarrow N$ between two manifolds M and N induces a linear *cotangent map* $T^*\varphi : T^*M \rightarrow T^*N$ between their cotangent bundles, i.e., the following diagram commutes:

$$\begin{array}{ccc} T^*M & \xrightarrow{T^*\varphi} & T^*N \\ \pi_M^* \downarrow & & \downarrow \pi_N^* \\ M & \xrightarrow{\varphi} & N \end{array}$$

All cotangent bundles and their cotangent maps form the category $T^*\mathcal{B}$. The category $T^*\mathcal{B}$ is the natural stage for *Hamiltonian dynamics*.

Now, we can formulate the *dual version of the global chain rule*. If $\varphi : M \rightarrow N$ and $\psi : N \rightarrow P$ are two smooth maps, then we have $T^*(\psi \circ \varphi) = T^*\psi \circ T^*\varphi$. In other words, we have a cofunctor $T^* : \mathcal{M} \Rightarrow T^*\mathcal{B}$ from the category \mathcal{M} of smooth manifolds to the category $T^*\mathcal{B}$ of their cotangent bundles:

$$\begin{array}{ccc} & M & \\ \varphi \swarrow & & \searrow (\psi \circ \varphi) \\ N & \xrightarrow{\psi} & P \end{array} \quad \xRightarrow{T^*} \quad \begin{array}{ccc} & T^*M & \\ T^*\varphi \swarrow & & \searrow T^*(\psi \circ \varphi) \\ T^*N & \xleftarrow{T^*\psi} & T^*P \end{array}$$

3.5.3 Application: Command/Control in Human-Robot Interactions

Suppose that we have a *human-robot team*, consisting of m robots and n humans. To be able to put the modelling of the fully controlled human-robot team performance into the rigorous geometrical settings, we suppose that all possible behaviors of m robots can be described by a set of continuous and smooth, time-dependent *robot configuration coordinates* $x^r = x^r(t)$, while all robot-related behaviors of n humans can be described by a set of

continuous and smooth, time-dependent *human configuration coordinates* $q^h = q^h(t)$. In other words, all robot coordinates, $x^r = x^r(t)$, constitute the smooth Riemannian manifold M_g^r (such that $r = 1, \dots, \dim(M_g^r)$), with the positive-definite metric form

$$g \mapsto ds^2 = g_{rs}(x)dx^r dx^s \quad (3.11)$$

similarly, all human coordinates $q^h = q^h(t)$, constitute a smooth Riemannian manifold N_a^h (such that $h = 1, \dots, \dim(N_a^h)$), with the positive-definite metric form

$$a \mapsto d\sigma^2 = a_{hk}(q)dq^h dq^k. \quad (3.12)$$

In this Riemannian geometry settings, the feedforward command/control action of humans upon robots is defined by a smooth map,

$$\mathcal{C} : N_a^h \rightarrow M_g^r,$$

which is in local coordinates given by a general (nonlinear) functional transformation

$$x^r = x^r(q^h), \quad (r = 1, \dots, \dim(M_g^r); h = 1, \dots, \dim(N_a^h)), \quad (3.13)$$

while its inverse, the feedback map from robots to humans is defined by a smooth map,

$$\mathcal{F} = \mathcal{C}^{-1} : M_g^r \rightarrow N_a^h,$$

which is in local coordinates given by an inverse functional transformation

$$q^h = q^h(x^r), \quad (h = 1, \dots, \dim(N_a^h); r = 1, \dots, \dim(M_g^r)). \quad (3.14)$$

Now, although the coordinate transformations (3.13) and (3.14) are completely general, nonlinear and even unknown at this stage, there is something known and simple about them: the corresponding transformations of differentials are *linear and homogenous*, namely

$$dx^r = \frac{\partial x^r}{\partial q^h} dq^h, \quad \text{and} \quad dq^h = \frac{\partial q^h}{\partial x^r} dx^r,$$

which imply linear and homogenous transformations of robot and human velocities,

$$\dot{x}^r = \frac{\partial x^r}{\partial q^h} \dot{q}^h, \quad \text{and} \quad \dot{q}^h = \frac{\partial q^h}{\partial x^r} \dot{x}^r. \quad (3.15)$$

Relation (3.15), representing two autonomous dynamical systems, given by two sets of ordinary differential equations (ODEs), geometrically defines two *velocity vector-fields*: (i) robot velocity vector-field, $v^r \equiv v^r(x^r, t) := \dot{x}^r(x^r, t)$; and human velocity vector-field, $u^h \equiv u^h(q^h, t) := \dot{q}^h(q^h, t)$. Recall that a vector-field defines a single vector at each point x^r (in some domain U) of a manifold in case. Its solution gives the *flow*, consisting of *integral curves* of the vector-field, such that all the vectors from the vector-field are tangent to integral curves at different points $x^i \in U$. Geometrically, a velocity vector-field is defined as a cross-section of the tangent bundle of the manifold. In our case, the robot velocity vector-field $v^r = \dot{x}^r(x^r, t)$ represents a cross-section of the robot tangent bundle TM_g^r , while the human velocity vector-field $u^h = \dot{q}^h(q^h, t)$ represents a cross-section of the human tangent bundle TN_a^h . In this way, two local velocity vector-fields, v^r and u^h , give local representations for the following two global *tangent maps*,

$$TC : TN_a^h \rightarrow TM_g^r, \quad \text{and} \quad T\mathcal{F} : TM_g^r \rightarrow TN_a^h.$$

To be able to proceed along the geometrodynamical line, we need next to formulate the two corresponding acceleration vector-fields, $a^r \equiv a^r(x^r, \dot{x}^r, t)$ and $w^h \equiv w^h(q^h, \dot{q}^h, t)$, as time rates of change of the two velocity vector-fields v^r and u^h . Now, recall that the acceleration vector-field is defined as the *absolute time derivative*, $\dot{v}^r = \frac{D}{dt}v^r$, of the velocity vector-field. In our case, we have the robotic acceleration vector-field $a^r := \dot{v}^r$ defined on M_g^r by

$$a^r := \dot{v}^r = \dot{v}^r + \Gamma_{st}^r v^s v^t = \ddot{x}^r + \Gamma_{st}^r \dot{x}^s \dot{x}^t, \quad (3.16)$$

and the human acceleration vector-field $w^h := \dot{u}^h$ defined on N_a^h by

$$w^h := \dot{u}^h = \dot{u}^h + \Gamma_{jk}^h u^j u^k = \ddot{q}^h + \Gamma_{jk}^h \dot{q}^j \dot{q}^k, \quad (3.17)$$

Geometrically, an acceleration vector-field is defined as a cross-section of the second tangent bundle of the manifold. In our case, the robot acceleration vector-field $a^r = \dot{v}^r(x^r, \dot{x}^r, t)$, given by the ODEs (3.16), represents a cross-section of the second robot tangent bundle TTM_g^r , while the human acceleration vector-field $w^h = \dot{u}^h(q^h, \dot{q}^h, t)$, given by the ODEs (3.17), represents a cross-section of the second human tangent bundle TTN_a^h . In this way, two local acceleration vector-fields, a^r and w^h , give local representations for the following two *second tangent maps*,

$$TTC : TTN_a^h \rightarrow TTM_g^r, \quad \text{and} \quad TTF : TTM_g^r \rightarrow TTN_a^h.$$

In other words, we have the feedforward command/control commutative diagram:

$$\begin{array}{ccc}
 TTN_a^h & \xrightarrow{TTC} & TTM_g^r \\
 \uparrow & & \uparrow \\
 TN_a^h & \xrightarrow{TC} & TM_g^r \\
 \uparrow & & \uparrow \\
 N_a^h & \xrightarrow{\mathcal{C}} & M_g^r
 \end{array}$$

as well as the feedback commutative diagram:

$$\begin{array}{ccc}
 TTN_a^h & \xleftarrow{TT\mathcal{F}} & TTM_g^r \\
 \uparrow & & \uparrow \\
 TN_a^h & \xleftarrow{T\mathcal{F}} & TM_g^r \\
 \uparrow & & \uparrow \\
 N_a^h & \xleftarrow{\mathcal{F}} & M_g^r
 \end{array}$$

These two commutative diagrams formally define the global feedforward and feedback human–robot interactions at the positional, velocity, and acceleration levels of command and control.

3.5.4 Application: *Generalized Bidirectional Associative Memory*

System Architecture

Here we present a covariant model of generalized bidirectional associative memory (GBAM), a neurodynamical classification machine generaliz-

ing Kosko BAM and RBAM systems (see [Kosko (1992)]). Mathematically, the GBAM is a tensor-field system (q, p, W) defined on a manifold M called the GBAM manifold. The system (q, p, W) includes two nonlinearly coupled (yet non-chaotic and stable) subsystems (see Figure 3.4): (i) activation (q, p) –dynamics, where q and p represent neuronal 1D tensor-fields, and (ii) self-organized learning W –dynamics, where W is a symmetric synaptic 2D tensor-field.

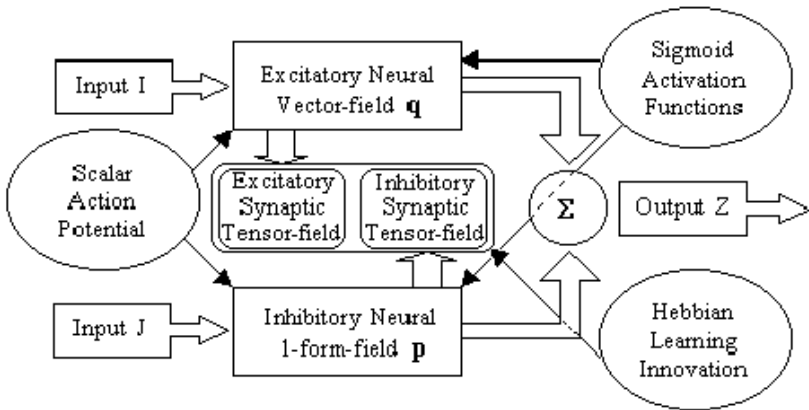


Fig. 3.4 Architecture of the GBAM neurodynamical classifier.

GBAM Activation Dynamics

The GBAM-manifold M can be viewed as a Banach space with a C^∞ –smooth structure on it, so that in each local chart U open in M , an nD smooth coordinate system U_α exists.

GBAM-activation (q, p) –dynamics, is defined as a system of two coupled, first-order oscillator tensor-fields, dual to each other, in a local Banach chart U_α , $\alpha = 1, \dots, n$ on M :

1. An excitatory neural vector-field $q^\alpha = q^\alpha(t) : M \rightarrow TM$, being a cross-section of the tangent bundle TM ; and
2. An inhibitory neural 1-form $p_\alpha = p_\alpha(t) : M \rightarrow T^*M$, being a cross-section of the cotangent bundle T^*M .

To start with conservative linear (q, p) –system, we postulate the GBAM

scalar activation–potential V to be a negative bilinear form:

$$V = -\frac{1}{2}\omega_{\alpha\beta}q^\alpha q^\beta - \frac{1}{2}\omega^{\alpha\beta}p_\alpha p_\beta + q^\alpha p_\alpha, \quad (\alpha, \beta = 1, \dots, n), \quad (3.18)$$

where n is the number of neurons in each neural field, while $\omega_{\alpha\beta}$ and $\omega^{\alpha\beta}$ represent respectively inhibitory–covariant and excitatory–contravariant components of the symmetric (with zero–trace) coupling GBAM synaptic tensor W .

The Lyapunov–stable, conservative, linear (q, \mathbf{p}) –dynamics is given as a bidirectional (excitatory–inhibitory) gradient system:

$$\dot{q}^\alpha = -\frac{\partial V}{\partial p_\alpha} = \omega^{\alpha\beta}p_\beta - q^\alpha, \quad \dot{p}_\alpha = -\frac{\partial V}{\partial q^\alpha} = \omega_{\alpha\beta}q^\beta - p_\alpha. \quad (3.19)$$

As W is a symmetric and zero–trace synaptic coupling tensor, the conservative linear dynamics (3.19) is equivalent to the rule that the state of each neuron (in both excitatory and inhibitory neural fields) is changed in time if and only if the scalar action potential V , defined by relation (3.18), is lowered. Therefore, the scalar action potential V is a monotonically non–increasing Lyapunov function $\dot{V} \leq 0$ for the conservative linear (q, p) –dynamics (3.19), which converges to a local minimum or ground state of V .

Applying the inputs I^α and J_α , we get the non–conservative linear (q, p) –system equations:

$$\dot{q}^\alpha = I^\alpha + \omega^{\alpha\beta}p_\beta - q^\alpha, \quad \dot{p}_\alpha = J_\alpha + \omega_{\alpha\beta}q^\beta - p_\alpha. \quad (3.20)$$

Further, applying the sigmoid GBAM activation functions $S_\alpha(\cdot)$ and $S^\alpha(\cdot)$ to the synaptic product–terms, we get the non-conservative nonlinear (q, p) –system equations, which generalize the transient RC–circuit neurodynamical model:

$$\dot{q}^\alpha = I^\alpha + \omega^{\alpha\beta}S_\beta(p_\beta) - q^\alpha, \quad \dot{p}_\alpha = J_\alpha + \omega_{\alpha\beta}S^\beta(q^\beta) - p_\alpha. \quad (3.21)$$

The equations in (3.21) represent a 2–input system that can be applied e.g., to classification of two–feature data. The generalization to an N –input system working in a ND *feature–space* is given by

$$\dot{q}_e^\alpha = I_e^\alpha + \omega_e^{\alpha\beta}S_\beta(p_\beta) - q_e^\alpha, \quad \dot{p}_\alpha^o = I_\alpha^o + \omega_{\alpha\beta}^oS^\beta(q^\beta) - p_\alpha^o, \quad (3.22)$$

where $e(= 2, 4, \dots, N)$ and $o(= 1, 3, \dots, N - 1)$ denote respectively even and odd partitions of the total sample of N features.

The GBAM model (3.22) gives a generalization of four well–known recurrent NN models:

1. Continuous Hopfield amplifier–circuit model [Hopfield (1984)]

$$C_j \dot{v}_j = I_j - \frac{v_j}{R_j} + T_{ij} u_i, \quad (i, j = 1, \dots, N),$$

where $v_j = v_j(t)$ represent the activation potentials in the j th processing unit, C_j and R_j denote input capacitances and leakage resistances, $u_i = f_i[v_i(t)]$ are output functions from processing elements, and $T_{ij} = w_{ij}$ is the inverse of the resistors connection–matrix; and the functions f_i are sigmoidal.

2. Cohen–Grossberg general ART–system [Cohen and Grossberg (1983)],

$$\dot{v}_j = -a_j(v_j)[b_j(v_j) - f_k(v_k)m_{jk}], \quad (j = 1, \dots, N),$$

with proved asymptotical stability.

3. Hecht–Nielsen counter–propagation network [Hecht–Nielsen (1987)],

$$\dot{v}_j = -Av_j + (B - v_j)I_j - v_j I_k,$$

where A, B are positive constants and I_j are input values for each processing unit.

4. Kosko's BAM (ABAM and RABAM) bidirectional models [Kosko (1992)]

$$\begin{aligned} \dot{v}_j &= -a_j(v_j)[b_j(v_j) - f_k(v_k)m_{jk}], \\ \dot{u}_k &= -a_k(u_k)[b_k(u_k) - f_j(u_j)m_{jk}], \end{aligned}$$

which is globally stable for the cases of signal and random–signal Hebbian learning.

GBAM Self–Organized Learning Dynamics

The continuous (and at least C^1 –differentiable) unsupervised update law for the coupling synaptic GBAM tensor–field W can be viewed both as an inhibitory–covariant Hebbian learning scheme, generalized from [Kosko (1992)]:

$$\dot{\omega}_{\alpha\beta} = -\omega_{\alpha\beta} + \Phi_{\alpha\beta}(q^\alpha, p_\alpha), \quad (\alpha, \beta = 1, \dots, n), \quad (3.23)$$

and, as an excitatory–contravariant Hebbian learning scheme:

$$\dot{\omega}^{\alpha\beta} = -\omega^{\alpha\beta} + \Phi^{\alpha\beta}(q^\alpha, p_\alpha), \quad (3.24)$$

where the three terms from the left to the right denote respectively the new–update value, the old value and the innovation of the synaptic tensor W . In this case the nonlinear (usually sigmoid) innovation functions $\Phi_{\alpha\beta}$ and $\Phi^{\alpha\beta}$ are defined by one of following four Hebbian models:

Signal Hebbian learning, with innovation in both variance–forms:

$$\begin{aligned}\Phi_{\alpha\beta} &= S_{\alpha}(q^{\alpha}) S_{\beta}(p_{\beta}), \\ \Phi^{\alpha\beta} &= S^{\alpha}(q^{\alpha}) S^{\beta}(p_{\beta});\end{aligned}\tag{3.25}$$

Differential Hebbian learning, with innovation in both variance–forms:

$$\begin{aligned}\Phi_{\alpha\beta} &= S_{\alpha}(q^{\alpha})S_{\beta}(p_{\beta}) + \dot{S}_{\alpha}(q^{\alpha})\dot{S}_{\beta}(p_{\beta}), \\ \Phi^{\alpha\beta} &= S^{\alpha}(q^{\alpha})S^{\beta}(p_{\beta}) + \dot{S}^{\alpha}(q^{\alpha})\dot{S}^{\beta}(p_{\beta}),\end{aligned}\tag{3.26}$$

where \dot{S} –terms denote the so–called ‘signal velocities’ (for details see [Kosko (1992)]).

Random signal Hebbian learning, with innovation in both variance–forms:

$$\begin{aligned}\Phi_{\alpha\beta} &= S_{\alpha}(q^{\alpha}) S_{\beta}(p_{\beta}) + n_{\alpha\beta}, \\ \Phi^{\alpha\beta} &= S^{\alpha}(q^{\alpha}) S^{\beta}(p_{\beta}) + n^{\alpha\beta},\end{aligned}\tag{3.27}$$

where $n_{\alpha\beta} = \{n_{\alpha\beta}(t)\}$, $n^{\alpha\beta} = \{n^{\alpha\beta}(t)\}$ respectively denote covariant and contravariant additive, zero–mean, Gaussian white–noise processes independent of the main innovation signal; and

Random differential signal Hebbian learning, with innovation in both variance–forms:

$$\begin{aligned}\Phi_{\alpha\beta} &= S_{\alpha}(q^{\alpha})S_{\beta}(p_{\beta}) + \dot{S}_{\alpha}(q^{\alpha})\dot{S}_{\beta}(p_{\beta}) + n_{\alpha\beta}, \\ \Phi^{\alpha\beta} &= S^{\alpha}(q^{\alpha})S^{\beta}(p_{\beta}) + \dot{S}^{\alpha}(q^{\alpha})\dot{S}^{\beta}(p_{\beta}) + n^{\alpha\beta}.\end{aligned}\tag{3.28}$$

Total GBAM (q, p, W) –neurodynamics and biological interpretation

Total GBAM tensorial neurodynamics is defined as a union of the neural oscillatory activation (q, p) –dynamics (3.22) and the synaptic learning

W —dynamics (3.28), namely

$$\begin{aligned}
 \dot{q}_e^\alpha &= I_e^\alpha + \omega_e^{\alpha\beta} S_\beta(p_\beta) - q_e^\alpha, \\
 \dot{p}_\alpha^o &= I_\alpha^o + \omega_{\alpha\beta}^o S^\beta(q^\beta) - p_\alpha^o, \\
 \dot{\omega}_e^{\alpha\beta} &= -\omega_e^{\alpha\beta} + \Phi_e^{\alpha\beta}(q^\alpha, p_\alpha), \\
 \dot{\omega}_{\alpha\beta}^o &= -\omega_{\alpha\beta}^o + \Phi_{\alpha\beta}^o(q^\alpha, p_\alpha), \quad (\alpha, \beta = 1, \dots, n),
 \end{aligned} \tag{3.29}$$

where the tensorial innovation Φ —functions are given by one of Hebbian models (3.25–3.28), $\alpha(= 1, \dots, n)$ is the number of continuous-time (or, graded-response) neurons in each neural-activation field, $e(= 2, 4, \dots, N)$ and $o(= 1, 3, \dots, N - 1)$ denote respectively even and odd partitions of the total sample of N features.

Artificial neural networks are generally inspired by biological neural systems, but in fact, some important features of biological systems are not present in most artificial neural networks. In particular, unidirectional neural networks, which include all associative neural networks except the BAM model introduced by [Kosko (1992)], do not resemble oscillatory biological neural systems. GBAM is a generalization of Kosko's ABAM and RABAM neural systems and inherits their oscillatory (excitatory/inhibitory) neuro-synaptic behavior. Such oscillatory behavior is a basic characteristic of a number of biological systems. Examples of similar oscillatory neural ensembles in the human nervous system are:

- Motoneurons and Renshaw interneurons in the spinal cord;
- Pyramidal and basket cells in the hippocampus;
- Mitral and granule cells in the olfactory bulb;
- Pyramidal cells and thalamic inter-neurons in cortico-thalamic system;
- Interacting excitatory and inhibitory populations of neurons found in the cerebellum, olfactory cortex, and neocortex, all representing the basic mechanisms for the generation of oscillating (EEG-monitored) activity in the brain.

Therefore, GBAM can be considered as a model for any of above-mentioned oscillatory biological neural systems.

3.6 Tensor Fields on Smooth Manifolds

3.6.1 Tensor Bundle

A *tensor bundle* \mathcal{T} associated to a smooth n -manifold M is defined as a *tensor product* of tangent and cotangent bundles:

$$\mathcal{T} = \bigotimes^q T^*M \otimes \bigotimes^p TM = TM \otimes \overbrace{\dots \otimes TM}^{p \text{ times}} \otimes T^*M \otimes \overbrace{\dots \otimes T^*M}^{q \text{ times}}.$$

Tensor bundles are special case of more general fibre bundles (see section 4.1 below).

A *tensor-field* of type (p, q) on a smooth n -manifold M is defined as a *smooth section* $\tau : M \rightarrow \mathcal{T}$ of the tensor bundle \mathcal{T} . The coefficients of the tensor-field τ are smooth (C^k) functions with p indices up and q indices down. The classical position of indices can be explained in modern terms as follows. If (U, ϕ) is a chart at a point $m \in M$ with local coordinates (x^1, \dots, x^n) , we have the *holonomous frame field*

$$\partial_{x^{i_1}} \otimes \partial_{x^{i_2}} \otimes \dots \otimes \partial_{x^{i_p}} \otimes dx^{j_1} \otimes dx^{j_2} \dots \otimes dx^{j_q},$$

for $i \in \{1, \dots, n\}^p$, $j = \{1, \dots, n\}^q$, over U of this tensor bundle, and for any (p, q) -tensor-field τ we have

$$\tau|U = \tau_{j_1 \dots j_q}^{i_1 \dots i_p} \partial_{x^{i_1}} \otimes \partial_{x^{i_2}} \otimes \dots \otimes \partial_{x^{i_p}} \otimes dx^{j_1} \otimes dx^{j_2} \dots \otimes dx^{j_q}.$$

For such tensor-fields the *Lie derivative* along any vector-field is defined (see section 3.7 below), and it is a *derivation* (i.e., both linearity and Leibniz rules hold) with respect to the tensor product. Tensor bundle \mathcal{T} admits many natural transformations (see [Kolar *et al.* (1993)]). For example, a ‘contraction’ like the trace $T^*M \otimes TM = L(TM, TM) \rightarrow M \times \mathbb{R}$, but applied just to one specified factor of type T^*M and another one of type TM , is a natural transformation. And any ‘permutation of the same kind of factors’ is a natural transformation.

The tangent bundle $\pi_M : TM \rightarrow M$ of a manifold M (introduced above) is a special tensor bundle over M such that, given an atlas $\{(U_\alpha, \varphi_\alpha)\}$ of M , TM has the *holonomic atlas*

$$\Psi = \{(U_\alpha, \varphi_\alpha = T\varphi_\alpha)\}.$$

The associated linear bundle coordinates are the induced coordinates (\dot{x}^λ) at a point $m \in M$ with respect to the *holonomic frames* $\{\partial_\lambda\}$ in tangent

spaces $T_m M$. Their transition functions read

$$\dot{x}'^\lambda = \frac{\partial x'^\lambda}{\partial x^\mu} \dot{x}^\mu.$$

Technically, the tangent bundle TM is a tensor bundle with the structure Lie group $GL(\dim M, \mathbb{R})$ (see section 3.8 below).

Recall that the cotangent bundle of M is the dual T^*M of TM . It is equipped with the induced coordinates (\dot{x}_λ) at a point $m \in M$ with respect to *holonomic coframes* $\{dx^\lambda\}$ dual of $\{\partial_\lambda\}$. Their transition functions read

$$\dot{x}'_\lambda = \frac{\partial x'^\mu}{\partial x^\lambda} \dot{x}_\mu.$$

3.6.1.1 Pull-Back and Push-Forward

In this section we define two important operations, following [Abraham *et al.* (1988)], which will be used in the further text.

Let $\varphi : M \rightarrow N$ be a C^k map of manifolds and $f \in C^k(N, \mathbb{R})$. Define the *pull-back* of f by φ by

$$\varphi^* f = f \circ \varphi \in C^k(M, \mathbb{R}).$$

If f is a C^k diffeomorphism and $X \in \mathcal{X}^k(M)$, the *push-forward* of X by φ is defined by

$$\varphi_* X = T\varphi \circ X \circ \varphi^{-1} \in \mathcal{X}^k(N).$$

If x^i are local coordinates on M and y^j local coordinates on N , the preceding formula gives the components of $\varphi_* X$ by

$$(\varphi_* X)^j(y) = \frac{\partial \varphi^j}{\partial x^i}(x) X^i(x), \quad \text{where} \quad y = \varphi(x).$$

We can interchange pull-back and push-forward by changing φ to φ^{-1} , that is, defining φ_* (resp. φ^*) by $\varphi_* = (\varphi^{-1})^*$ (resp. $\varphi^* = (\varphi^{-1})_*$). Thus the *push-forward* of a function f on M is $\varphi_* f = f \circ \varphi^{-1}$ and the *pull-back* of a vector-field Y on N is $\varphi^* Y = (T\varphi)^{-1} \circ Y \circ \varphi$.

Notice that φ must be a diffeomorphism in order that the pull-back and push-forward operations make sense, the only exception being pull-back of functions. Thus vector-fields can only be pulled back and pushed forward by diffeomorphisms. However, even when φ is not a diffeomorphism we can talk about *φ -related vector-fields* as follows.

Let $\varphi : M \rightarrow N$ be a C^k map of manifolds. The vector-fields $X \in \mathcal{X}^{k-1}(M)$ and $Y \in \mathcal{X}^{k-1}(N)$ are called φ -related, denoted $X \sim_\varphi Y$, if $T\varphi \circ X = Y \circ \varphi$.

Note that if φ is diffeomorphism and X and Y are φ -related, then $Y = \varphi_* X$. However, in general, X can be φ -related to more than one vector-field on N . φ -relatedness means that the following diagram commutes:

$$\begin{array}{ccc} TM & \xrightarrow{T\varphi} & TN \\ X \uparrow & & \uparrow Y \\ M & \xrightarrow{\varphi} & N \end{array}$$

The behavior of flows under these operations is as follows: Let $\varphi : M \rightarrow N$ be a C^k -map of manifolds, $X \in \mathcal{X}^k(M)$ and $Y \in \mathcal{X}^k(N)$. Let F_t and G_t denote the flows of X and Y respectively. Then $X \sim_\varphi Y$ iff $\varphi \circ F_t = G_t \circ \varphi$. In particular, if φ is a diffeomorphism, then the equality $Y = \varphi_* X$ holds iff the flow of Y is $\varphi \circ F_t \circ \varphi^{-1}$ (This is called the push-forward of F_t by φ since it is the natural way to construct a diffeomorphism on N out of one on M). In particular, $(F_t)_* X = X$. Therefore, the flow of the push-forward of a vector-field is the push-forward of its flow.

3.6.1.2 Dynamical Evolution and Flow

As a motivational example, consider a mechanical system that is capable of assuming various states described by points in a set U . For example, U might be $\mathbb{R}^3 \times \mathbb{R}^3$ and a state might be the positions and momenta (x^i, p_i) of a particle moving under the influence of the central force field, with $i = 1, 2, 3$. As time passes, the state evolves. If the state is $\gamma_0 \in U$ at time s and this changes to γ at a later time t , we set

$$F_{t,s}(\gamma_0) = \gamma,$$

and call $F_{t,s}$ the *evolution operator*; it maps a state at time s to what the state would be at time t ; that is, after time $t - s$ has elapsed. Determinism is expressed by the *Chapman-Kolmogorov law* [Abraham *et al.* (1988)]:

$$F_{\tau,t} \circ F_{t,s} = F_{\tau,s}, \quad F_{t,t} = \text{identity}. \quad (3.30)$$

The evolution laws are called *time independent*, or *autonomous*, when $F_{t,s}$ depends only on $t - s$. In this case the preceding law (3.30) becomes

the *group property*:

$$F_t \circ F_s = F_{t+s}, \quad F_0 = \text{identity}. \quad (3.31)$$

We call such an F_t a *flow* and $F_{t,s}$ a *time-dependent flow*, or an evolution operator. If the system is irreversible, that is, defined only for $t \geq s$, we speak of a *semi-flow* [Abraham *et al.* (1988)].

Usually, instead of $F_{t,s}$ the *laws of motion* are given in the form of ODEs that we must solve to find the flow. These equations of motion have the form:

$$\dot{\gamma} = X(\gamma), \quad \gamma(0) = \gamma_0,$$

where X is a (possibly time-dependent) vector-field on U .

As a continuation of the previous example, consider the motion of a particle of mass m under the influence of the central force field (like gravity, or Coulombic potential) F^i ($i = 1, 2, 3$), described by the *Newtonian equation of motion*:

$$m\ddot{x}^i = F^i(x). \quad (3.32)$$

By introducing momenta $p_i = m\dot{x}^i$, equation (3.32) splits into two Hamiltonian equations:

$$\dot{x}^i = p_i/m, \quad \dot{p}_i = F_i(x). \quad (3.33)$$

Note that in Euclidean space we can freely interchange subscripts and superscripts. However, in general case of a Riemannian manifold, $p_i = mg_{ij}\dot{x}^j$ and (3.33) properly reads

$$\dot{x}^i = g^{ij}p_j/m, \quad \dot{p}_i = F_i(x). \quad (3.34)$$

The phase-space here is the *Riemannian manifold* $(\mathbb{R}^3 \setminus \{0\}) \times \mathbb{R}^3$, that is, the cotangent bundle of $\mathbb{R}^3 \setminus \{0\}$, which is itself a smooth manifold for the central force field. The r.h.s of equations (3.34) define a Hamiltonian vector-field on this 6D manifold by

$$X(x, p) = ((x^i, p_i), (p_i/m, F_i(x))). \quad (3.35)$$

Integration of equations (3.34) produces trajectories (in this particular case, planar conic sections). These trajectories comprise the flow F_t of the vector-field $X(x, p)$ defined in (3.35).

3.6.1.3 Vector-Fields and Their Flows

3.6.1.4 Vector-Fields on M

A *vector-field* X on U , where U is an open chart in n -manifold M , is a *smooth function* from U to M assigning to each point $m \in U$ a vector at that point, i.e., $X(m) = (m, X(m))$. If $X(m)$ is tangent to M for each $m \in M$, X is said to be a *tangent vector-field* on M . If $X(m)$ is orthogonal to M (i.e., $X(p) \in M_m^\perp$) for each $X(m) \in M$, X is said to be a *normal vector-field* on M .

In other words, let M be a C^k -manifold. A C^k -vector-field on M is a C^k -section of the tangent bundle TM of M . Thus a vector-field X on a manifold M is a C^k -map $X : M \rightarrow TM$ such that $X(m) \in T_m M$ for all points $m \in M$, and $\pi_M \circ X = Id_M$. Therefore, a vector-field assigns to each point m of M a vector based (i.e., bound) at that point. The set of all C^k vector-fields on M is denoted by $\mathcal{X}^k(M)$.

A vector-field $X \in \mathcal{X}^k(M)$ represents a field of direction indicators [Thirring (1979)]: to every point m of M it assigns a vector in the tangent space $T_m M$ at that point. If X is a vector-field on M and (U, ϕ) is a chart on M and $m \in U$, then we have $X(m) = X(m) \phi^i \frac{\partial}{\partial \phi^i}$. Following [Kolar *et al.* (1993)], we write $X|_U = X \phi^i \frac{\partial}{\partial \phi^i}$.

Let M be a connected n -manifold, and let $f : U \rightarrow \mathbb{R}$ (U an open set in M) and $c \in \mathbb{R}$ be such that $M = f^{-1}(c)$ (i.e., M is the *level set* of the function f at *height* c) and $\nabla f(m) \neq 0$ for all $m \in M$. Then there exist on M exactly two smooth unit normal vector-fields $N_{1,2}(m) = \pm \frac{\nabla f(m)}{|\nabla f(m)|}$ (here $|X| = (X \cdot X)^{1/2}$ denotes the norm or length of a vector X , and (\cdot) denotes the scalar product on M) for all $m \in M$, called *orientations* on M .

Let $\varphi : M \rightarrow N$ be a smooth map. Recall that two vector-fields $X \in \mathcal{X}^k(M)$ and $Y \in \mathcal{X}(N)$ are called φ -related, if $T\varphi \circ X = Y \circ \varphi$ holds, i.e., if the following diagram commutes:

$$\begin{array}{ccc}
 TM & \xrightarrow{T\varphi} & TN \\
 \uparrow X & & \uparrow Y \\
 M & \xrightarrow{\varphi} & N
 \end{array}$$

In particular, a diffeomorphism $\varphi : M \rightarrow N$ induces a linear map between vector-fields on two manifolds, $\varphi^* : \mathcal{X}^k(M) \rightarrow \mathcal{X}(N)$, such that $\varphi^* X =$

$T\varphi \circ X \circ \varphi^{-1} : N \rightarrow TN$, i.e., the following diagram commutes:

$$\begin{array}{ccc}
 TM & \xrightarrow{T\varphi} & TN \\
 X \uparrow & & \uparrow \varphi^* X \\
 M & \xrightarrow{\varphi} & N
 \end{array}$$

The correspondences $M \rightarrow TM$ and $\varphi \rightarrow T\varphi$ obviously define a functor $T : \mathcal{M} \Rightarrow \mathcal{M}$ from the category of smooth manifolds to itself. T is another special case of the vector bundle functor (4.3.2), and closely related to the tangent bundle functor (3.5).

A C^k *time-dependent vector-field* is a C^k -map $X : \mathbb{R} \times M \rightarrow TM$ such that $X(t, m) \in T_m M$ for all $(t, m) \in \mathbb{R} \times M$, i.e., $X_t(m) = X(t, m)$.

3.6.1.5 Integral Curves as Dynamical Trajectories

Recall (3.5) that a curve γ at a point m of an n -manifold M is a C^0 -map from an open interval $I \subset \mathbb{R}$ into M such that $0 \in I$ and $\gamma(0) = m$. For such a curve we may assign a tangent vector at each point $\gamma(t)$, $t \in I$, by $\dot{\gamma}(t) = T_t \gamma(1)$.

Let X be a smooth tangent vector-field on the smooth n -manifold M , and let $m \in M$. Then there exists an open interval $I \subset \mathbb{R}$ containing 0 and a parameterized curve $\gamma : I \rightarrow M$ such that:

- (1) $\gamma(0) = m$;
- (2) $\dot{\gamma}(t) = X(\gamma(t))$ for all $t \in I$; and
- (3) If $\beta : \tilde{I} \rightarrow M$ is any other parameterized curve in M satisfying (1) and (2), then $\tilde{I} \subset I$ and $\beta(t) = \gamma(t)$ for all $t \in \tilde{I}$.

A parameterized curve $\gamma : I \rightarrow M$ satisfying condition (2) is called an *integral curve* of the tangent vector-field X . The unique γ satisfying conditions (1)–(3) is the *maximal integral curve* of X through $m \in M$.

In other words, let $\gamma : I \rightarrow M$, $t \mapsto \gamma(t)$ be a smooth curve in a manifold M defined on an interval $I \subseteq \mathbb{R}$. $\dot{\gamma}(t) = \frac{d}{dt} \gamma(t)$ defines a smooth vector-field along γ since we have $\pi_M \circ \dot{\gamma} = \gamma$. Curve γ is called an *integral curve* or *flow line* of a vector-field $X \in \mathcal{X}^k(M)$ if the tangent vector determined by γ equals X at every point $m \in M$, i.e.,

$$\dot{\gamma} = X \circ \gamma,$$

or, if the following diagram commutes:

$$\begin{array}{ccc}
 TI & \xrightarrow{Tu} & TM \\
 \uparrow 1 & \nearrow \dot{\gamma} & \uparrow X \\
 I & \xrightarrow{\gamma} & M
 \end{array}$$

On a chart (U, ϕ) with coordinates $\phi(m) = (x^1(m), \dots, x^n(m))$, for which $\varphi \circ \gamma : t \mapsto \gamma_i(t)$ and $T\varphi \circ X \circ \varphi^{-1} : x^i \mapsto (x^i, X_i(m))$, this is written

$$\dot{\gamma}_i(t) = X_i(\gamma(t)), \text{ for all } t \in I \subseteq \mathbb{R}, \quad (3.36)$$

which is an ordinary differential equation of first-order in n dimensions.

The *velocity* $\dot{\gamma}$ of the parameterized curve $\gamma(t)$ is a vector-field along γ defined by

$$\dot{\gamma}(t) = (\gamma(t), \dot{x}^1(t), \dots, \dot{x}^n(t)).$$

Its length $|\dot{\gamma}| : I \rightarrow \mathbb{R}$, defined by $|\dot{\gamma}|(t) = |\dot{\gamma}(t)|$ for all $t \in I$, is a function along α . $|\dot{\gamma}|$ is called *speed* of γ [Arnold (1989)].

Each vector-field X along γ is of the form $X(t) = (\gamma(t), X_1(t), \dots, X_n(t))$, where each component X_i is a function along γ . X is *smooth* if each $X_i : I \rightarrow M$ is smooth. The *derivative* of a smooth vector-field X along a curve $\gamma(t)$ is the vector-field \dot{X} along γ defined by

$$\dot{X}(t) = (\gamma(t), \dot{X}_1(t), \dots, \dot{X}_n(t)).$$

$\dot{X}(t)$ measures the *rate of change of the vector part* $(X_1(t), \dots, X_n(t))$ of $X(t)$ along γ . Thus, the *acceleration* $\ddot{\gamma}(t)$ of a parameterized curve $\gamma(t)$ is the vector-field along γ get by differentiating the velocity field $\dot{\gamma}(t)$.

Differentiation of vector-fields along parameterized curves has the following properties. For X and Y smooth vector-fields on M along the parameterized curve $\gamma : I \rightarrow M$ and f a smooth function along γ , we have:

- (1) $\frac{d}{dt}(X + Y) = \dot{X} + \dot{Y}$;
- (2) $\frac{d}{dt}(fX) = \dot{f}X + f\dot{X}$; and
- (3) $\frac{d}{dt}(X \cdot Y) = \dot{X}Y + X\dot{Y}$.

A *geodesic* in M is a parameterized curve $\gamma : I \rightarrow M$ whose acceleration $\ddot{\gamma}$ is everywhere orthogonal to M ; that is, $\ddot{\gamma}(t) \in M_{\alpha(t)}^\perp$ for all $t \in I \subset \mathbb{R}$. Thus a geodesic is a curve in M which always goes ‘straight ahead’ in the surface. Its acceleration serves only to keep it in the surface. It has no component of acceleration tangent to the surface. Therefore, it also has a constant speed $\dot{\gamma}(t)$.

Let $v \in M_m$ be a vector on M . Then there exists an open interval $I \subset \mathbb{R}$ containing 0 and a geodesic $\gamma : I \rightarrow M$ such that:

- (1) $\gamma(0) = m$ and $\dot{\gamma}(0) = v$; and
- (2) If $\beta : \tilde{I} \rightarrow M$ is any other geodesic in M with $\beta(0) = m$ and $\dot{\beta}(0) = v$, then $\tilde{I} \subset I$ and $\beta(t) = \gamma(t)$ for all $t \in \tilde{I}$.

The geodesic γ is now called the *maximal geodesic* in M passing through m with initial velocity v .

By definition, a parameterized curve $\gamma : I \rightarrow M$ is a geodesic of M iff its acceleration is everywhere perpendicular to M , i.e., iff $\ddot{\gamma}(t)$ is a multiple of the orientation $N(\gamma(t))$ for all $t \in I$, i.e., $\ddot{\gamma}(t) = g(t)N(\gamma(t))$, where $g : I \rightarrow \mathbb{R}$. Taking the scalar product of both sides of this equation with $N(\gamma(t))$ we find $g = -\dot{\gamma} \cdot \dot{N}(\gamma(t))$. Thus $\gamma : I \rightarrow M$ is geodesic iff it satisfies the differential equation

$$\ddot{\gamma}(t) + \dot{N}(\gamma(t)) \cdot \dot{\gamma}(t) = 0.$$

This vector equation represents the system of second-order component ODEs

$$\ddot{x}^i + N_i(x + 1, \dots, x^n) \frac{\partial N_j}{\partial x^k}(x + 1, \dots, x^n) \dot{x}^j \dot{x}^k = 0.$$

The substitution $u^i = \dot{x}^i$ reduces this second-order differential system (in n variables x^i) to the first-order differential system

$$\dot{x}^i = u^i, \quad \dot{u}^i = -N_i(x + 1, \dots, x^n) \frac{\partial N_j}{\partial x^k}(x + 1, \dots, x^n) \dot{x}^j \dot{x}^k$$

(in $2n$ variables x^i and u^i). This first-order system is just the differential equation for the integral curves of the vector-field X in $U \times \mathbb{R}$ (U open chart in M), in which case X is called a *geodesic spray*.

Now, when an integral curve $\gamma(t)$ is the path a mechanical system Ξ follows, i.e., the solution of the equations of motion, it is called a *trajectory*. In this case the parameter t represents time, so that (3.36) describes motion of the system Ξ on its configuration manifold M .

If $X_i(m)$ is C^0 the existence of a local solution is guaranteed, and a *Lipschitz condition* would imply that it is unique. Therefore, exactly one integral curve passes through every point, and different integral curves can never cross. As $X \in \mathcal{X}^k(M)$ is C^k , the following statement about the solution with arbitrary initial conditions holds [Thirring (1979); Arnold (1989)]:

Theorem. Given a vector-field $X \in \mathcal{X}(M)$, for all points $p \in M$, there exist $\eta > 0$, a neighborhood V of p , and a function $\gamma : (-\eta, \eta) \times V \rightarrow M$, $(t, x^i(0)) \mapsto \gamma(t, x^i(0))$ such that

$$\dot{\gamma} = X \circ \gamma, \quad \gamma(0, x^i(0)) = x^i(0) \quad \text{for all } x^i(0) \in V \subseteq M.$$

For all $|t| < \eta$, the map $x^i(0) \mapsto \gamma(t, x^i(0))$ is a diffeomorphism f_t^X between V and some open set of M . For proof, see [Dieudonne (1969)], I, 10.7.4 and 10.8.

This Theorem states that trajectories that are near neighbors cannot suddenly be separated. There is a well-known estimate (see [Dieudonne (1969)], I, 10.5) according to which points cannot diverge faster than exponentially in time if the derivative of X is uniformly bounded.

An integral curve $\gamma(t)$ is said to be *maximal* if it is not a restriction of an integral curve defined on a larger interval $I \subseteq \mathbb{R}$. It follows from the existence and uniqueness theorems for ODEs with smooth r.h.s and from elementary properties of Hausdorff spaces that for any point $m \in M$ there exists a maximal integral curve γ_m of X , passing for $t = 0$ through point m , i.e., $\gamma(0) = m$.

Theorem (Local Existence, Uniqueness, and Smoothness) [Abraham *et al.* (1988)]. Let E be a Banach space, $U \subset E$ be open, and suppose $X : U \subset E \rightarrow E$ is of class C^k , $k \geq 1$. Then

1. For each $x_0 \in U$, there is a curve $\gamma : I \rightarrow U$ at x_0 such that $\dot{\gamma}(t) = X(\gamma(t))$ for all $t \in I$.
2. Any two such curves are equal on the intersection of their domains.
3. There is a neighborhood U_0 of the point $x_0 \in U$, a real number $a > 0$, and a C^k map $F : U_0 \times I \rightarrow E$, where I is the open interval $] -a, a[$, such that the curve $\gamma_u : I \rightarrow E$, defined by $\gamma_u(t) = F(u, t)$ is a curve at $u \in E$ satisfying the ODEs $\dot{\gamma}_u(t) = X(\gamma_u(t))$ for all $t \in I$.

Proposition (Global Uniqueness). Suppose γ_1 and γ_2 are two integral curves of a vector-field X at a point $m \in M$. Then $\gamma_1 = \gamma_2$ on the intersection of their domains [Abraham *et al.* (1988)].

If for every point $m \in M$ the curve γ_m is defined on the entire real axis

\mathbb{R} , then the vector-field X is said to be *complete*.

The *support* of a vector-field X defined on a manifold M is defined to be the closure of the set $\{m \in M | X(m) = 0\}$. A C^k vector-field with compact support on a manifold M is complete. In particular, a C^k vector-field on a compact manifold is complete. Completeness corresponds to well-defined dynamics persisting eternally.

Now, following [Abraham *et al.* (1988)], for the *derivative* of a C^k function $f : E \rightarrow \mathbb{R}$ in the direction X we use the notation $X[f] = df \cdot X$, where df stands for the *derivative map*. In standard coordinates on \mathbb{R}^n this is a standard *gradient*

$$df(x) = \nabla f = (\partial_{x^1} f, \dots, \partial_{x^n} f), \quad \text{and} \quad X[f] = X^i \partial_{x^i} f.$$

Let F_t be the flow of X . Then $f(F_t(x)) = f(F_s(x))$ if $t \geq s$.

For example, Newtonian equations for a moving particle of mass m in a potential field V in \mathbb{R}^n are given by $\ddot{q}^i(t) = -(1/m)\nabla V(q^i(t))$, for a smooth function $V : \mathbb{R}^n \rightarrow \mathbb{R}$. If there are constants $a, b \in \mathbb{R}$, $b \geq 0$ such that $(1/m)V(q^i) \geq a - b\|q^i\|^2$, then *every solution exists for all time*. To show this, rewrite the second-order equations as a first-order system $\dot{q}^i = (1/m)p_i$, $\dot{p}_i = -V(q^i)$ and note that the energy $E(q^i, p_i) = (1/2m)\|p_i\|^2 + V(q)$ is a *first integral* of the motion. Thus, for any solution $(q^i(t), p_i(t))$ we have $E(q^i(t), p_i(t)) = E(q^i(0), p_i(0)) = V(q(0))$.

Let X_t be a C^k time-dependent vector-field on an n -manifold M , $k \geq 1$, and let m_0 be an *equilibrium* of X_t , that is, $X_t(m_0) = 0$ for all t . Then for any T there exists a neighborhood V of m_0 such that any $m \in V$ has integral curve existing for time $t \in [-T, T]$.

3.6.1.6 Dynamical Flows on M

Recall (6.289) that the *flow* F_t of a C^k vector-field $X \in \mathcal{X}^k(M)$ is the *one-parameter group of diffeomorphisms* $F_t : M \rightarrow M$ such that $t \mapsto F_t(m)$ is the integral curve of X with initial condition m for all $m \in M$ and $t \in I \subseteq \mathbb{R}$. The flow $F_t(m)$ is C^k by induction on k . It is defined as [Abraham *et al.* (1988)]:

$$\frac{d}{dt} F_t(x) = X(F_t(x)).$$

Existence and uniqueness theorems for ODEs guarantee that F_t is smooth in m and t . From uniqueness, we get the *flow property*:

$$F_{t+s} = F_t \circ F_s$$

along with the initial conditions $F_0 = \text{identity}$. The flow property generalizes the situation where $M = V$ is a linear space, $X(x) = Ax$ for a (bounded) linear operator A , and where $F_t(x) = e^{tA}x$ – to the nonlinear case. Therefore, the flow $F_t(m)$ can be defined as a *formal exponential*

$$F_t(m) = \exp(tX) = (I + tX + \frac{t^2}{2}X^2 + \dots) = \sum_{k=0}^{\infty} \frac{X^k t^k}{k!}.$$

recall that a *time-dependent vector-field* is a map $X : M \times \mathbb{R} \rightarrow TM$ such that $X(m, t) \in T_m M$ for each point $m \in M$ and $t \in \mathbb{R}$. An *integral curve* of X is a curve $\gamma(t)$ in M such that

$$\dot{\gamma}(t) = X(\gamma(t), t), \quad \text{for all } t \in I \subseteq \mathbb{R}.$$

In this case, the flow is the one-parameter group of diffeomorphisms $F_{t,s} : M \rightarrow M$ such that $t \mapsto F_{t,s}(m)$ is the integral curve $\gamma(t)$ with initial condition $\gamma(s) = m$ at $t = s$. Again, the existence and uniqueness Theorem from ODE-theory applies here, and in particular, uniqueness gives the time-dependent flow property, i.e., the *Chapman–Kolmogorov law*

$$F_{t,r} = F_{t,s} \circ F_{s,r}.$$

If X happens to be time independent, the two notions of flows are related by $F_{t,s} = F_{t-s}$ (see [Marsden and Ratiu (1999)]).

3.6.1.7 Categories of ODEs

Ordinary differential equations are naturally organized into their categories (see [Kock (1981)]). First order ODEs are organized into a category ODE_1 . A first-order ODE on a manifold-like object M is a vector-field $X : M \rightarrow TM$, and a *morphism of vector-fields* $(M_1, X_1) \rightarrow (M_2, X_2)$ is a map $f : M_1 \rightarrow M_2$ such that the following diagram commutes

$$\begin{array}{ccc} TM_1 & \xrightarrow{Tf} & TM_2 \\ \uparrow X_1 & & \uparrow X_2 \\ M_1 & \xrightarrow{f} & M_2 \end{array}$$

A *global solution* of the differential equation (M, X) , or a *flow line* of a vector-field X , is a morphism from $(\mathbb{R}, \frac{\partial}{\partial x})$ to (M, X) .

Similarly, second-order ODEs are organized into a category ODE_2 . A second-order ODE on M is usually constructed as a vector-field on TM , $\xi : TM \rightarrow TTM$, and a morphism of vector-fields $(M_1, \xi_1) \rightarrow (M_2, \xi_2)$ is a map $f : M_1 \rightarrow M_2$ such that the following diagram commutes

$$\begin{array}{ccc} TTM_1 & \xrightarrow{TTf} & TTM_2 \\ \uparrow \xi_1 & & \uparrow \xi_2 \\ TM_1 & \xrightarrow{Tf} & TM_2 \end{array}$$

Unlike solutions for first-order ODEs, solutions for second-order ODEs are not in general homomorphisms from \mathbb{R} , unless the second-order ODE is a *spray* [Kock and Reyes (2003)].

3.6.2 Differential Forms on Smooth Manifolds

Recall (see section 2.1.4.2 above) that *exterior differential forms* are a special kind of antisymmetrical covariant tensors, that formally occur as *integrands* under ordinary integral signs in \mathbb{R}^3 . To give a more precise exposition, here we start with 1-forms, which are dual to vector-fields, and after that introduce general k -forms.

3.6.2.1 1-Forms on M

Dual to the notion of a C^k vector-field X on an n -manifold M is a C^k *covector-field*, or a C^k 1-form α , which is defined as a C^k -section of the cotangent bundle T^*M , i.e., $\alpha : M \rightarrow T^*M$ is smooth and $\pi_M^* \circ \alpha = Id_M$. We denote the set of all C^k 1-forms by $\Omega^1(M)$. A basic example of a 1-form is the differential df of a real-valued function $f \in C^k(M, \mathbb{R})$. With point wise addition and scalar multiplication $\Omega^1(M)$ becomes a vector space.

In other words, a C^k 1-form α on a C^k manifold M is a real-valued function on the set of all tangent vectors to M , i.e., $\alpha : TM \rightarrow \mathbb{R}$ with the following properties:

- (1) α is linear on the tangent space $T_m M$ for each $m \in M$;
- (2) For any C^k vector-field $X \in \mathcal{X}^k(M)$, the function $f : M \rightarrow \mathbb{R}$ is C^k .

Given a 1-form α , for each point $m \in M$ the map $\alpha(m) : T_m M \rightarrow \mathbb{R}$ is an element of the dual space $T_m^* M$. Therefore, the space of 1-forms $\Omega^1(M)$

is dual to the space of vector-fields $\mathcal{X}^k(M)$.

In particular, the *coordinate 1-forms* dx^1, \dots, dx^n are locally defined at any point $m \in M$ by the property that for any vector-field $X = (X^1, \dots, X^n) \in \mathcal{X}^k(M)$,

$$dx^i(X) = X^i.$$

The dx^i 's form a *basis* for the 1-forms at any point $m \in M$, with local coordinates (x^1, \dots, x^n) , so any 1-form α may be expressed in the form

$$\alpha = f_i(m) dx^i.$$

If a vector-field X on M has the form $X(m) = (X^1(m), \dots, X^n(m))$, then at any point $m \in M$,

$$\alpha_m(X) = f_i(m) X^i(m),$$

where $f \in C^k(M, \mathbb{R})$.

Suppose we have a 1D closed curve $\gamma = \gamma(t)$ inside a smooth manifold M . Using a simplified ‘physical’ notation, a 1-form $\alpha(x)$ defined at a point $x \in M$, given by

$$\alpha(x) = \alpha_i(x) dx^i, \quad (3.37)$$

can be unambiguously integrated over a curve $\gamma \in M$, as follows. Parameterize γ by a parameter t , so that its coordinates are given by $x^i(t)$. At time t , the velocity $\dot{x} = \dot{x}(t)$ is a tangent vector to M at $x(t)$. One can insert this tangent vector into the linear map $\alpha(x)$ to get a real number. By definition, inserting the vector $\dot{x}(t)$ into the linear map dx^i gives the component $\dot{x}^i = \dot{x}^i(t)$. Doing this for every t , we can then integrate over t ,

$$\int (\alpha_i(x(t)) \dot{x}^i) dt. \quad (3.38)$$

Note that this expression is independent of the parametrization in terms of t . Moreover, from the way that tangent vectors transform, one can deduce how the linear maps dx^i should transform, and from this how the coefficients $\alpha_i(x)$ should transform. Doing this, one sees that the above expression is also invariant under changes of coordinates on M . Therefore, a 1-form can be unambiguously integrated over a curve in M . We write such an integral as

$$\int_{\gamma} \alpha_i(x) dx^i, \quad \text{or, even shorter, as} \quad \int_{\gamma} \alpha.$$

Clearly, when M is itself a 1D manifold, (3.38) gives precisely the ordinary integration of a function $\alpha(x)$ over x , so the above notation is indeed natural.

The 1-forms on M are part of an algebra, called the *exterior algebra*, or *Grassmann algebra* on M . The multiplication \wedge in this algebra is called *wedge product* (see (3.40) below), and it is skew-symmetric,

$$dx^i \wedge dx^j = -dx^j \wedge dx^i.$$

One consequence of this is that $dx^i \wedge dx^i = 0$.

3.6.2.2 k -Forms on M

A differential form, or an exterior form α of degree k , or a k -form for short, is a section of the vector bundle $\Lambda^k T^*M$, i.e., $\alpha : M \rightarrow \Lambda^k T^*M$. In other words, $\alpha(m) : T_m M \times \dots \times T_m M \rightarrow \mathbb{R}$ (with k factors $T_m M$) is a function that assigns to each point $m \in M$ a skew-symmetric k -multilinear map on the tangent space $T_m M$ to M at m . Without the skew-symmetry assumption, α would be called a $(0, k)$ -tensor-field. The space of all k -forms is denoted by $\Omega^k(M)$. It may also be viewed as the space of all skew symmetric $(0, k)$ -tensor-fields, the space of all maps

$$\Phi : \mathcal{X}^k(M) \times \dots \times \mathcal{X}^k(M) \rightarrow C^k(M, \mathbb{R}),$$

which are k -linear and skew-symmetric (see (3.40) below). We put $\Omega^k(M) = C^k(M, \mathbb{R})$.

In particular, a 2-form ω on an n -manifold M is a section of the vector bundle $\Lambda^2 T^*M$. If (U, ϕ) is a chart at a point $m \in M$ with local coordinates (x^1, \dots, x^n) let $\{e_1, \dots, e_n\} = \{\partial_{x^1}, \dots, \partial_{x^n}\}$ – be the corresponding basis for $T_m M$, and let $\{e^1, \dots, e^n\} = \{dx^1, \dots, dx^n\}$ – be the dual basis for $T_m^* M$. Then at each point $m \in M$, we can write a 2-form ω as

$$\omega_m(v, u) = \omega_{ij}(m) v^i u^j, \quad \text{where } \omega_{ij}(m) = \omega_m(\partial_{x^i}, \partial_{x^j}).$$

Similarly to the case of a 1-form α (3.37), one would like to define a 2-form ω as something which can naturally be integrated over a 2D surface Σ within a smooth manifold M . At a specific point $x \in M$, the tangent plane to such a surface is spanned by a pair of tangent vectors, (\dot{x}^1, \dot{x}^2) . So, to generalize the construction of a 1-form, we should give a bilinear map from such a pair to \mathbb{R} . The most general form of such a map is

$$\omega_{ij}(x) dx^i \otimes dx^j, \tag{3.39}$$

where the tensor product of two cotangent vectors acts on a pair of vectors as,

$$dx^i \otimes dx^j (\dot{x}^1, \dot{x}^2) = dx^i(\dot{x}^1) dx^j(\dot{x}^2).$$

On the r.h.s. of this equation, one multiplies two ordinary numbers got by letting the linear map dx^i act on \dot{x}^1 , and dx^j on \dot{x}^2 .

However, the bilinear map (3.39) is slightly too general to give a good integration procedure. The reason is that we would like the integral to change sign if we change the orientation of integration, just like in the 1D case. In 2D, changing the orientation means exchanging \dot{x}^1 and \dot{x}^2 , so we want our bilinear map to be antisymmetric under this exchange. This is achieved by defining a 2-form to be

$$\omega = \omega_{ij}(x) (dx^i \otimes dx^j - dx^j \otimes dx^i) \equiv \omega_{ij}(x) dx^i \wedge dx^j$$

We now see why a 2-form corresponds to an *antisymmetric* tensor-field: the symmetric part of ω_{ij} would give a vanishing contribution to ω . Now, parameterizing a surface Σ in M with two coordinates t_1 and t_2 , and reasoning exactly like we did in the case of a 1-form, one can show that the integration of a 2-form over such a surface is indeed well-defined, and independent of the parametrization of both Σ and M .

If each summand of a differential form $\alpha \in \Omega^k(M)$ contains k basis 1-forms dx^i 's, the form is called a k -form. Functions $f \in C^k(M, \mathbb{R})$ are considered to be 0-forms, and any form on an n -manifold M of degree $k > n$ must be zero due to the skew-symmetry.

Any k -form $\alpha \in \Omega^k(M)$ may be expressed in the form

$$\alpha = f_I dx^{i_1} \wedge \dots \wedge dx^{i_k} = f_I dx^I,$$

where I is a *multiindex* $I = (i_1, \dots, i_k)$ of length k , and \wedge is the wedge product which is associative, bilinear and anticommutative.

Just as 1-forms act on vector-fields to give real-valued functions, so k -forms act on k -tuples of vector-fields to give real-valued functions.

The *wedge product* of two differential forms, a k -form $\alpha \in \Omega^k(M)$ and an l -form $\beta \in \Omega^l(M)$ is a $(k+l)$ -form $\alpha \wedge \beta$ defined as:

$$\alpha \wedge \beta = \frac{(k+l)!}{k!l!} \mathbf{A}(\alpha \otimes \beta), \quad (3.40)$$

where $\mathbf{A} : \Omega^k(M) \times \Omega^l(M) \rightarrow \Omega^{k+l}(M)$, $\mathbf{A}\tau(e_1, \dots, e_k) = \frac{1}{k!} \sum_{\sigma \in S_k} (\text{sign } \sigma) \tau(e_{\sigma(1)}, \dots, e_{\sigma(k)})$, where S_k is the permutation group on k elements consisting of all bijections $\sigma : \{1, \dots, k\} \rightarrow \{1, \dots, k\}$.

For any k -form $\alpha \in \Omega^k(M)$ and l -form $\beta \in \Omega^l(M)$, the wedge product is defined fiberwise, i.e., $(\alpha \wedge \beta)_m = \alpha_x \wedge \beta_m$ for each point $m \in M$. It is also associative, i.e., $(\alpha \wedge \beta) \wedge \gamma = \alpha \wedge (\beta \wedge \gamma)$, and graded commutative, i.e., $\alpha \wedge \beta = (-1)^{kl} \beta \wedge \alpha$. These properties are proved in multilinear algebra. So $M \Rightarrow \Omega^k(M)$ is a contravariant functor from the category \mathcal{M} into the category of real graded commutative algebras [Kolar *et al.* (1993)].

Let M be an n -manifold, $X \in \mathcal{X}^k(M)$, and $\alpha \in \Omega^{k+1}(M)$. The *interior product*, or *contraction*, $i_X \alpha = X \lrcorner \alpha \in \Omega^k(M)$ of X and α (with *insertion operator* i_X) is defined as

$$i_X \alpha(X^1, \dots, X^k) = \alpha(X, X^1, \dots, X^k).$$

Insertion operator i_X of a vector-field $X \in \mathcal{X}^k(M)$ is natural with respect to the pull-back F^* of a diffeomorphism $F : M \rightarrow N$ between two manifolds, i.e., the following diagram commutes:

$$\begin{array}{ccc} \Omega^k(N) & \xrightarrow{F^*} & \Omega^k(M) \\ i_X \downarrow & & \downarrow i_{F^*X} \\ \Omega^{k-1}(N) & \xrightarrow{F^*} & \Omega^{k-1}(M) \end{array}$$

Similarly, insertion operator i_X of a vector-field $X \in \mathcal{X}^k(M)$ is natural with respect to the push-forward F_* of a diffeomorphism $F : M \rightarrow N$, i.e., the following diagram commutes:

$$\begin{array}{ccc} \Omega^k(M) & \xrightarrow{F_*} & \Omega^k(N) \\ i_Y \downarrow & & \downarrow i_{F_*Y} \\ \Omega^{k-1}(M) & \xrightarrow{F_*} & \Omega^{k-1}(N) \end{array}$$

In case of *Riemannian manifolds* there is another exterior operation. Let M be a smooth n -manifold with Riemannian metric $g = \langle, \rangle$ and the corresponding volume element μ . The *Hodge star operator* $*$: $\Omega^k(M) \rightarrow \Omega^{n-k}(M)$ on M is defined as

$$\alpha \wedge * \beta = \langle \alpha, \beta \rangle \mu \quad \text{for } \alpha, \beta \in \Omega^k(M).$$

The Hodge star operator satisfies the following properties for $\alpha, \beta \in \Omega^k(M)$ [Abraham *et al.* (1988)]:

- (1) $\alpha \wedge * \beta = \langle \alpha, \beta \rangle \mu = \beta \wedge * \alpha$;
- (2) $*1 = \mu, \quad * \mu = (-1)^{Ind(g)}$;
- (3) $* * \alpha = (-1)^{Ind(g)} (-1)^{k(n-k)} \alpha$;
- (4) $\langle \alpha, \beta \rangle = (-1)^{Ind(g)} \langle * \alpha, * \beta \rangle$, where $Ind(g)$ is the *index* of the metric g .

3.6.2.3 Exterior Differential Systems

Here we give an informal introduction to *exterior differential systems* (EDS, for short), which are expressions involving differential forms related to any manifold M . Later, when we fully develop the necessary differential geometrical as well as variational machinery (see (5.8) below), we will give a more precise definition of EDS.

Central in the language of EDS is the notion of *coframing*, which is a real finite-dimensional smooth manifold M with a given global *cobasis* and coordinates, but without requirement for a proper topological and differential structures. For example, $M = \mathbb{R}^3$ is a coframing with cobasis $\{dx, dy, dz\}$ and coordinates $\{x, y, z\}$. In addition to the cobasis and coordinates, a coframing can be given structure equations (3.10.2.4) and restrictions. For example, $M = \mathbb{R}^2 \setminus \{0\}$ is a coframing with cobasis $\{e^1, e^2\}$, a single coordinate $\{r\}$, structure equations $\{dr = e^1, de^1 = 0, de^2 = e^1 \wedge e^2 / r\}$ and restrictions $\{r \neq 0\}$.

A *system* S on M in EDS terminology is a list of expressions including differential forms (e.g., $S = \{dz - ydx\}$).

Now, a simple EDS is a triple (S, Ω, M) , where S is a system on M , and Ω is an *independence condition*: either a decomposable k -form or a system of k -forms on M . An EDS is a list of simple EDS objects where the various coframings are all disjoint.

An *integral element* of an exterior system (S, Ω, M) is a subspace $P \subset T_m M$ of the tangent space at some point $m \in M$ such that all forms in S vanish when evaluated on vectors from P . Alternatively, an integral element $P \subset T_m M$ can be represented by its annihilator $P^\perp \subset T_m^* M$, comprising those 1-forms at m which annul every vector in P . For example, with $M = \mathbb{R}^3 = \{(x, y, z)\}$, $S = \{dx \wedge dz\}$ and $\Omega = \{dx, dz\}$, the integral element $P = \{\partial_x + \partial_z, \partial_y\}$ is equally determined by its annihilator $P^\perp = \{dz - dx\}$. Again, for $S = \{dz - ydx\}$ and $\Omega = \{dx\}$, the integral element $P = \{\partial_x + y\partial_z\}$ can be specified as $\{dy\}$.

3.6.3 Exterior Derivative and (Co)Homology

The *exterior derivative* is an operation that takes k -forms to $(k+1)$ -forms on a smooth manifold M . It defines a unique family of maps $d : \Omega^k(U) \rightarrow \Omega^{k+1}(U)$, U open in M , such that (see [Abraham *et al.* (1988)]):

- (1) d is a \wedge -*antiderivation*; that is, d is \mathbb{R} -linear and for two forms $\alpha \in \Omega^k(U)$, $\beta \in \Omega^l(U)$,

$$d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^k \alpha \wedge d\beta.$$

- (2) If $f \in C^k(U, \mathbb{R})$ is a function on M , then $df = \frac{\partial f}{\partial x^i} dx^i : M \rightarrow T^*M$ is the differential of f , such that $df(X) = i_X df = \mathcal{L}_X f - di_X f = \mathcal{L}_X f = X[f]$ for any $X \in \mathcal{X}^k(M)$.
- (3) $d^2 = d \circ d = 0$ (that is, $d^{k+1}(U) \circ d^k(U) = 0$).
- (4) d is natural with respect to restrictions $|_U$; that is, if $U \subset V \subset M$ are open and $\alpha \in \Omega^k(V)$, then $d(\alpha|_U) = (d\alpha)|_U$, or the following diagram commutes:

$$\begin{array}{ccc} \Omega^k(V) & \xrightarrow{|_U} & \Omega^k(U) \\ d \downarrow & & \downarrow d \\ \Omega^{k+1}(V) & \xrightarrow{|_U} & \Omega^{k+1}(U) \end{array}$$

- (5) d is natural with respect to the Lie derivative \mathcal{L}_X (4.3.2) along any vector-field $X \in \mathcal{X}^k(M)$; that is, for $\omega \in \Omega^k(M)$ we have $\mathcal{L}_X \omega \in \Omega^k(M)$ and $d\mathcal{L}_X \omega = \mathcal{L}_X d\omega$, or the following diagram commutes:

$$\begin{array}{ccc} \Omega^k(M) & \xrightarrow{\mathcal{L}_X} & \Omega^k(M) \\ d \downarrow & & \downarrow d \\ \Omega^{k+1}(M) & \xrightarrow{\mathcal{L}_X} & \Omega^{k+1}(M) \end{array}$$

- (6) Let $\varphi : M \rightarrow N$ be a C^k map of manifolds. Then $\varphi^* : \Omega^k(N) \rightarrow \Omega^k(M)$ is a homomorphism of differential algebras (with \wedge and d) and d is natural with respect to $\varphi^* = F^*$; that is, $\varphi^* d\omega = d\varphi^* \omega$, or the following

diagram commutes:

$$\begin{array}{ccc} \Omega^k(N) & \xrightarrow{\varphi^*} & \Omega^k(M) \\ d \downarrow & & \downarrow d \\ \Omega^{k+1}(N) & \xrightarrow{\varphi^*} & \Omega^{k+1}(M) \end{array}$$

- (7) Analogously, d is natural with respect to diffeomorphism $\varphi_* = (F^*)^{-1}$; that is, $\varphi_* d\omega = d\varphi_* \omega$, or the following diagram commutes:

$$\begin{array}{ccc} \Omega^k(N) & \xrightarrow{\varphi_*} & \Omega^k(M) \\ d \downarrow & & \downarrow d \\ \Omega^{k+1}(N) & \xrightarrow{\varphi_*} & \Omega^{k+1}(M) \end{array}$$

- (8) $\mathcal{L}_X = i_X \circ d + d \circ i_X$ for any $X \in \mathcal{X}^k(M)$ (the Cartan ‘magic’ formula).
 (9) $\mathcal{L}_X \circ d = d \circ \mathcal{L}_X$, i.e., $[\mathcal{L}_X, d] = 0$ for any $X \in \mathcal{X}^k(M)$.
 (10) $[\mathcal{L}_X, i_Y] = i_{[X, Y]}$; in particular, $i_X \circ \mathcal{L}_X = \mathcal{L}_X \circ i_X$ for all $X, Y \in \mathcal{X}^k(M)$.

Given a k -form $\alpha = f_I dx^I \in \Omega^k(M)$, the exterior derivative is defined in local coordinates (x^1, \dots, x^n) of a point $m \in M$ as

$$d\alpha = d(f_I dx^I) = \frac{\partial f_I}{\partial x^{i_k}} dx^{i_k} \wedge dx^I = df_I \wedge dx^{i_1} \wedge \dots \wedge dx^{i_k}.$$

In particular, the exterior derivative of a function $f \in C^k(M, \mathbb{R})$ is a 1-form $df \in \Omega^1(M)$, with the property that for any $m \in M$, and $X \in \mathcal{X}^k(M)$,

$$df_m(X) = X(f),$$

i.e., $df_m(X)$ is a Lie derivative of f at m in the direction of X . Therefore, in local coordinates (x^1, \dots, x^n) of a point $m \in M$ we have

$$df = \frac{\partial f}{\partial x^i} dx^i.$$

For any two functions $f, g \in C^k(M, \mathbb{R})$, exterior derivative obeys the *Leibniz rule*:

$$d(fg) = g df + f dg,$$

and the *chain rule*:

$$d(g(f)) = g'(f) df.$$

A k -form $\alpha \in \Omega^k(M)$ is called *closed form* if $d\alpha = 0$, and it is called *exact form* if there exists a $(k-1)$ -form $\beta \in \Omega^{k-1}(M)$ such that $\alpha = d\beta$. Since $d^2 = 0$, every exact form is closed. The converse is only partially true (Poincaré Lemma): every closed form is *locally exact*. This means that given a closed k -form $\alpha \in \Omega^k(M)$ on an open set $U \subset M$, any point $m \in U$ has a neighborhood on which there exists a $(k-1)$ -form $\beta \in \Omega^{k-1}(U)$ such that $d\beta = \alpha|_U$.

The Poincaré lemma is a generalization and unification of two well-known facts in vector calculus:

- (1) If $\text{curl } F = 0$, then locally $F = \text{grad } f$;
- (2) If $\text{div } F = 0$, then locally $F = \text{curl } G$.

Poincaré lemma for contractible manifolds: Any closed form on a smoothly contractible manifold is exact.

3.6.3.1 Intuition Behind Cohomology

The simple formula $d^2 = 0$ leads to the important topological notion of *cohomology*. Let us try to solve the equation $d\omega = 0$ for a p -form ω . A trivial solution is $\omega = 0$. From the above formula, we can actually find a much larger class of trivial solutions: $\omega = d\alpha$ for a $(p-1)$ -form α . More generally, if ω is any solution to $d\omega = 0$, then so is $\omega + d\alpha$. We want to consider these two solutions as equivalent:

$$\omega \sim \omega + \omega' \quad \text{if} \quad \omega' \in \text{Im } d,$$

where $\text{Im } d$ is the image of d , that is, the collection of all p -forms of the form $d\alpha$. (To be precise, the image of d contains q -forms for any $0 < q \leq n$, so we should restrict this image to the p -forms for the p we are interested in.) The set of all p -forms which satisfy $d\omega = 0$ is called the kernel of d , denoted $\text{Ker } d$, so we are interested in $\text{Ker } d$ up to the *equivalence classes* defined by adding elements of $\text{Im } d$. (Again, strictly speaking, $\text{Ker } d$ consists of q -forms for several values of q , so we should restrict it to the p -forms for our particular choice of p .) This set of equivalence classes is called $H^p(M)$, the p -th *de Rham cohomology group* of M ,

$$H^p(M) = \frac{\text{Ker } d}{\text{Im } d}.$$

Clearly, $\text{Ker } d$ is a group under addition: if two forms $\omega^{(1)}$ and $\omega^{(2)}$ satisfy $d\omega^{(1)} = d\omega^{(2)} = 0$, then so does $\omega^{(1)} + \omega^{(2)}$. Moreover, if we change $\omega^{(i)}$ by adding some $d\alpha^{(i)}$, the result of the addition will still be in the same cohomology class, since it differs from $\omega^{(1)} + \omega^{(2)}$ by $d(\alpha^{(1)} + \alpha^{(2)})$. Therefore, we can view this addition really as an addition of cohomology classes: $H^p(M)$ is itself an additive group. Also note that if $\omega^{(3)}$ and $\omega^{(4)}$ are in the same cohomology class (that is, their difference is of the form $d\alpha^{(3)}$), then so are $c\omega^{(3)}$ and $c\omega^{(4)}$ for any constant factor c . In other words, we can multiply a cohomology class by a constant to get another cohomology class: cohomology classes actually form a *vector space*.

3.6.3.2 Intuition Behind Homology

Another operator similar to the exterior derivative d is the *boundary operator* δ , which maps compact submanifolds of a smooth manifold M to their boundary. Here, $\delta C = 0$ means that a submanifold C of M has no boundary, and $C = \delta U$ means that C is itself the boundary of some submanifold U . It is intuitively clear, and not very hard to prove, that $\delta^2 = 0$: the boundary of a compact submanifold does not have a boundary itself. That the objects on which δ acts are independent of its coordinates is also clear. So is the grading of the objects: the degree p is the dimension of the submanifold C .³ What is less clear is that the collection of submanifolds actually forms a vector space, but one can always *define* this vector space to consist of formal linear combinations of submanifolds, and this is precisely how one proceeds. The p D elements of this vector space are called *p -chains*. One should think of $-C$ as C with its orientation reversed, and of the sum of two disjoint sets, $C^1 + C^2$, as their union. The equivalence classes constructed from δ are called *homology classes*.

For example, in Figure 3.5, C^1 and C^2 both satisfy $\delta C = 0$, so they are elements of $\text{Ker } \delta$. Moreover, it is clear that neither of them separately can be viewed as the boundary of another submanifold, so they are not in the trivial homology class $\text{Im } \delta$. However, the boundary of U is $C^1 - C^2$. (The minus sign in front of C^2 is a result of the fact that C^2 itself actually has the wrong orientation to be considered a boundary of U .) This can be written as $C^1 - C^2 = \delta U$, or equivalently $C^1 = C^2 + \delta U$, showing that C^1 and C^2 are in the same homology class.

The cohomology groups for the δ -operator are called *homology groups*,

³Note that here we have an example of an operator that maps objects of degree p to objects of degree $p - 1$ instead of $p + 1$.

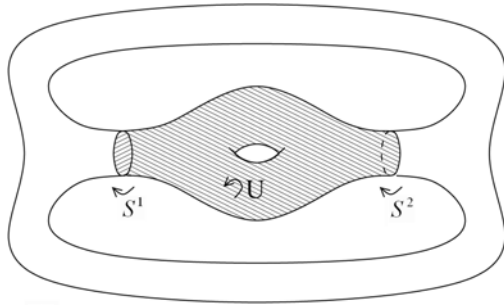


Fig. 3.5 The 1D submanifolds S^1 and S^2 represent the same homology class, since their difference is the boundary of U .

and denoted by $H_p(M)$, with a lower index.⁴ The p -chains C that satisfy $\delta C = 0$ are called p -cycles. Again, the $H_p(M)$ only exist for $0 \leq p \leq n$.

There is an interesting relation between cohomology and homology groups. Note that we can construct a bilinear map from $H^p(M) \times H_p(M) \rightarrow \mathbb{R}$ by

$$([\omega], [C]) \mapsto \int_C \omega, \quad (3.41)$$

where $[\omega]$ denotes the cohomology class of a p -form ω , and $[\Sigma]$ the homology class of a p -cycle Σ . Using Stokes' Theorem, it can be seen that the result does not depend on the representatives for either ω or C

$$\begin{aligned} \int_{C+\delta U} \omega + d\alpha &= \int_C \omega + \int_C d\alpha + \int_{\delta U} \omega + d\alpha \\ &= \int_C \omega + \int_{\delta C} \alpha + \int_U d(\omega + d\alpha) = \int_C \omega, \end{aligned}$$

where we used that by the definition of (co)homology classes, $\delta C = 0$ and $d\omega = 0$. As a result, the above map is indeed well-defined on homology and cohomology classes. A very important Theorem by de Rham says that this map is nondegenerate [De Rham (1984)]. This means that if we take some $[\omega]$ and we know the result of the map (3.41) for all $[C]$, this uniquely determines $[\omega]$, and similarly if we start by picking an $[C]$. This in particular means that the vector space $H^p(M)$ is the dual vector space of $H_p(M)$.

⁴Historically, as can be seen from the terminology, homology came first and cohomology was related to it in the way we will discuss below. However, since the cohomology groups have a more natural additive structure, it is the name 'cohomology' which is actually used for generalizations.

3.6.3.3 De Rham Complex and Homotopy Operators

After an intuitive introduction of (co)homology ideas, we now turn to their proper definitions. Given a smooth manifold M , let $\Omega^p(M)$ denote the space of all smooth p -forms on M . The differential d , mapping p -forms to $(p+1)$ -forms, serves to define the *de Rham complex* on M

$$0 \rightarrow \Omega^0(M) \xrightarrow{d^0} \Omega^1(M) \xrightarrow{d^1} \dots \xrightarrow{d^{n-1}} \Omega^n(M) \rightarrow 0. \quad (3.42)$$

Recall (from section 6.200 above) that in general, a *complex* is defined as a sequence of vector spaces, and linear maps between successive spaces, with the property that the composition of any pair of successive maps is identically 0. In the case of the de Rham complex (3.42), this requirement is a restatement of the closure property for the exterior differential: $d \circ d = 0$.

In particular, for $n = 3$, the de Rham complex on a manifold M reads

$$0 \rightarrow \Omega^0(M) \xrightarrow{d^0} \Omega^1(M) \xrightarrow{d^1} \Omega^2(M) \xrightarrow{d^2} \Omega^3(M) \rightarrow 0. \quad (3.43)$$

If $\omega \equiv f(x, y, z) \in \Omega^0(M)$, then

$$d^0 \omega \equiv d^0 f = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz = \text{grad } \omega.$$

If $\omega \equiv f dx + g dy + h dz \in \Omega^1(M)$, then

$$d^1 \omega \equiv \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y} \right) dx \wedge dy + \left(\frac{\partial h}{\partial y} - \frac{\partial g}{\partial z} \right) dy \wedge dz + \left(\frac{\partial f}{\partial z} - \frac{\partial h}{\partial x} \right) dz \wedge dx = \text{curl } \omega.$$

If $\omega \equiv F dy \wedge dz + G dz \wedge dx + H dx \wedge dy \in \Omega^2(M)$, then

$$d^2 \omega \equiv \frac{\partial F}{\partial x} + \frac{\partial G}{\partial y} + \frac{\partial H}{\partial z} = \text{div } \omega.$$

Therefore, the de Rham complex (3.43) can be written as

$$0 \rightarrow \Omega^0(M) \xrightarrow{\text{grad}} \Omega^1(M) \xrightarrow{\text{curl}} \Omega^2(M) \xrightarrow{\text{div}} \Omega^3(M) \rightarrow 0.$$

Using the closure property for the exterior differential, $d \circ d = 0$, we get the standard identities from vector calculus

$$\text{curl} \cdot \text{grad} = 0 \quad \text{and} \quad \text{div} \cdot \text{curl} = 0.$$

The definition of the complex requires that the kernel of one of the linear maps contains the image of the preceding map. The complex is

exact if this containment is *equality*. In the case of the de Rham complex (3.42), exactness means that a closed p -form ω , meaning that $d\omega = 0$, is necessarily an exact p -form, meaning that there exists a $(p-1)$ -form θ such that $\omega = d\theta$. (For $p = 0$, it says that a smooth function f is closed, $df = 0$, iff it is constant). Clearly, any exact form is closed, but the converse need not hold. Thus the de Rham complex is *not* in general exact. The celebrated *de Rham Theorem* states that the extent to which this complex fails to be exact measures purely topological information about the manifold M , its cohomology group.

On the local side, for special types of domains in Euclidean space \mathbb{R}^m , there is only trivial topology and we do have exactness of the de Rham complex (3.42). This result, known as the *Poincaré lemma*, holds for *star-shaped* domains $M \subset \mathbb{R}^m$: Let $M \subset \mathbb{R}^m$ be a star-shaped domain. Then the de Rham complex over M is exact.

The key to the proof of exactness of the de Rham complex lies in the construction of suitable *homotopy operators*. By definition, these are linear operators $h : \Omega^p \rightarrow \Omega^{p-1}$, taking differential p -forms into $(p-1)$ -forms, and satisfying the basic identity [Olver (1986)]

$$\omega = dh(\omega) + h(d\omega), \quad (3.44)$$

for all p -forms $\omega \in \Omega^p$. The discovery of such a set of operators immediately implies exactness of the complex. For if ω is closed, $d\omega = 0$, then (3.44) reduces to $\omega = d\theta$ where $\theta = h(\omega)$, so ω is exact.

3.6.3.4 Stokes Theorem and de Rham Cohomology

Stokes Theorem states that if α is an $(n-1)$ -form on an orientable n -manifold M , then the integral of $d\alpha$ over M equals the integral of α over ∂M , the boundary of M . The classical theorems of Gauss, Green, and Stokes are special cases of this result.

A *manifold with boundary* is a set M together with an atlas of charts (U, ϕ) with boundary on M . Define (see [Abraham *et al.* (1988)]) the *interior* and *boundary* of M respectively as

$$\text{Int } M = \bigcup_U \phi^{-1}(\text{Int}(\phi(U))), \quad \text{and} \quad \partial M = \bigcup_U \phi^{-1}(\partial(\phi(U))).$$

If M is a manifold with boundary, then its interior $\text{Int } M$ and its boundary ∂M are smooth manifolds without boundary. Moreover, if $f : M \rightarrow N$ is a diffeomorphism, N being another manifold with boundary, then f in-

duces, by restriction, two diffeomorphisms

$$\text{Int } f : \text{Int } M \rightarrow \text{Int } N, \quad \text{and} \quad \partial f : \partial M \rightarrow \partial N.$$

If $n = \dim M$, then $\dim(\text{Int } M) = n$ and $\dim(\partial M) = n - 1$.

To integrate a differential n -form over an n -manifold M , M must be oriented. If $\text{Int } M$ is oriented, we want to choose an *orientation* on ∂M compatible with it. As for manifolds without boundary a volume form on an n -manifold with boundary M is a nowhere vanishing n -form on M . Fix an orientation on \mathbb{R}^n . Then a chart (U, ϕ) is called *positively oriented* if the map $T_m \phi : T_m M \rightarrow \mathbb{R}^n$ is orientation preserving for all $m \in U$.

Let M be a compact, oriented k D smooth manifold with boundary ∂M . Let α be a smooth $(k - 1)$ -form on M . Then the classical *Stokes formula* holds

$$\int_M d\alpha = \int_{\partial M} \alpha.$$

If $\partial M = \emptyset$ then $\int_M d\alpha = 0$.

The quotient space

$$H^k(M) = \frac{\text{Ker}(d : \Omega^k(M) \rightarrow \Omega^{k+1}(M))}{\text{Im}(d : \Omega^{k-1}(M) \rightarrow \Omega^k(M))}$$

represents the k th *de Rham cohomology group* of a manifold M . recall that the *de Rham Theorem* states that these Abelian groups are isomorphic to the so-called singular cohomology groups of M defined in algebraic topology in terms of simplices and that depend only on the topological structure of M and not on its differentiable structure. The isomorphism is provided by integration; the fact that the integration map drops to the preceding quotient is guaranteed by Stokes' Theorem.

The exterior derivative commutes with the pull-back of differential forms. That means that the vector bundle $\Lambda^k T^* M$ is in fact the value of a functor, which associates a bundle over M to each manifold M and a vector bundle homomorphism over φ to each (local) diffeomorphism φ between manifolds of the same dimension. This is a simple example of the concept of a natural bundle. The fact that the exterior derivative d transforms sections of $\Lambda^k T^* M$ into sections of $\Lambda^{k+1} T^* M$ for every manifold M can be expressed by saying that d is an operator from $\Lambda^k T^* M$ into $\Lambda^{k+1} T^* M$. That the exterior derivative d commutes with (local) diffeomorphisms now means, that d is a natural operator from the functor $\Lambda^k T^*$ into functor $\Lambda^{k+1} T^*$. If $k > 0$, one can show that d is the unique

natural operator between these two natural bundles up to a constant. So even linearity is a consequence of naturality [Kolar *et al.* (1993)].

3.6.3.5 Euler–Poincaré Characteristics of M

The *Euler–Poincaré characteristics* of a manifold M equals the sum of its *Betti numbers*

$$\chi(M) = \sum_{p=0}^n (-1)^p b_p.$$

In case of $2n$ D oriented compact Riemannian manifold M (*Gauss–Bonnet Theorem*) its Euler–Poincaré characteristics is equal

$$\chi(M) = \int_M \gamma,$$

where γ is a closed $2n$ form on M , given by

$$\gamma = \frac{(-1)^n}{(4\pi)^n n!} \epsilon_{i_1 \dots i_{2n}} \Omega_{i_2}^{i_1} \wedge \Omega_{i_{2n}}^{i_{2n-1}},$$

where Ω_j^i is the curvature 2–form of a Riemannian connection on M .

Poincaré–Hopf Theorem: The Euler–Poincaré characteristics $\chi(M)$ of a compact manifold M equals the sum of indices of zeros of any vector–field on M which has only isolated zeros.

3.6.3.6 Duality of Chains and Forms on M

In topology of finite–dimensional smooth (i.e., C^{p+1} with $p \geq 0$) manifolds, a fundamental notion is the *duality* between p –chains C and p –forms (i.e., p –cochains) ω on the smooth manifold M , or domains of integration and integrands – as an integral on M represents a bilinear functional (see [Choquet–Bruhat and DeWitt–Morete (1982); Dodson and Parker (1997)])

$$\int_C \omega \equiv \langle C, \omega \rangle, \quad (3.45)$$

where the integral is called the *period* of ω . Period depends only on the cohomology class of ω and the homology class of C . A closed form (cocycle) is exact (coboundary) if all its periods vanish, i.e., $d\omega = 0$ implies $\omega = d\theta$. The duality (3.45) is based on the classical Stokes formula

$$\int_C d\omega = \int_{\partial C} \omega.$$

This is written in terms of scalar products on M as

$$\langle C, d\omega \rangle = \langle \partial C, \omega \rangle,$$

where ∂C is the boundary of the p -chain C oriented coherently with C . While the boundary operator ∂ is a global operator, the coboundary operator, that is, the exterior derivative d , is local, and thus more suitable for applications. The main property of the exterior differential,

$$d^2 = 0 \quad \text{implies} \quad \partial^2 = 0,$$

can be easily proved by the use of Stokes' formula

$$\langle \partial^2 C, \omega \rangle = \langle \partial C, d\omega \rangle = \langle C, d^2 \omega \rangle = 0.$$

The analysis of p -chains and p -forms on the finite-dimensional smooth manifold M is usually performed in (co)homology categories (see [Dodson and Parker (1997); Dieudonne (1988)]) related to M .

Let \mathcal{M}^\bullet denote the category of cochains, (i.e., p -forms) on the smooth manifold M . When $\mathcal{C} = \mathcal{M}^\bullet$, we have the category $\mathcal{S}^\bullet(\mathcal{M}^\bullet)$ of generalized cochain complexes A^\bullet in \mathcal{M}^\bullet , and if $A' = 0$ for $n < 0$ we have a subcategory $\mathcal{S}_{\mathcal{R}}^\bullet(\mathcal{M}^\bullet)$ of the de Rham differential complexes in \mathcal{M}^\bullet

$$\begin{aligned} A_{DR}^\bullet : 0 \rightarrow \Omega^0(M) &\xrightarrow{d} \Omega^1(M) \xrightarrow{d} \Omega^2(M) \cdots \\ \cdots &\xrightarrow{d} \Omega^n(M) \xrightarrow{d} \cdots \end{aligned} \quad (3.46)$$

Here $A' = \Omega^n(M)$ is the vector space over \mathbb{R} of all p -forms ω on M (for $p = 0$ the smooth functions on M) and $d_n = d : \Omega^{n-1}(M) \rightarrow \Omega^n(M)$ is the exterior differential. A form $\omega \in \Omega^n(M)$ such that $d\omega = 0$ is a closed form or n -cocycle. A form $\omega \in \Omega^n(M)$ such that $\omega = d\theta$, where $\theta \in \Omega^{n-1}(M)$, is an exact form or n -coboundary. Let $Z^n(M) = \text{Ker}(d)$ (resp. $B^n(M) = \text{Im}(d)$) denote a real vector space of cocycles (resp. coboundaries) of degree n . Since $d_{n+1} d_n = d^2 = 0$, we have $B^n(M) \subset Z^n(M)$. The quotient vector space

$$H_{DR}^n(M) = \text{Ker}(d) / \text{Im}(d) = Z^n(M) / B^n(M)$$

is the de Rham cohomology group. The elements of $H_{DR}^n(M)$ represent equivalence sets of cocycles. Two cocycles ω_1, ω_2 belong to the same equivalence set, or are cohomologous (written $\omega_1 \sim \omega_2$) iff they differ by a coboundary $\omega_1 - \omega_2 = d\theta$. The de Rham cohomology class of any form $\omega \in \Omega^n(M)$ is $[\omega] \in H_{DR}^n(M)$. The de Rham differential complex (3.46) can

be considered as a system of second-order ODEs $d^2\theta = 0$, $\theta \in \Omega^{n-1}(M)$ having a solution represented by $Z^n(M) = \text{Ker}(d)$.

Analogously let \mathcal{M}_\bullet denote the category of chains on the smooth manifold M . When $\mathcal{C} = \mathcal{M}_\bullet$, we have the category $\mathcal{S}_\bullet(\mathcal{M}_\bullet)$ of generalized chain complexes A_\bullet in \mathcal{M}_\bullet , and if $A_n = 0$ for $n < 0$ we have a subcategory $\mathcal{S}_\bullet^{\mathcal{C}}(\mathcal{M}_\bullet)$ of chain complexes in \mathcal{M}_\bullet .

$$A_\bullet : 0 \leftarrow C^0(M) \xleftarrow{\partial} C^1(M) \xleftarrow{\partial} C^2(M) \cdots \xleftarrow{\partial} C^n(M) \xleftarrow{\partial} \cdots.$$

Here $A_n = C^n(M)$ is the vector space over \mathbb{R} of all finite chains C on the manifold M and $\partial_n = \partial : C^{n+1}(M) \rightarrow C^n(M)$. A finite chain C such that $\partial C = 0$ is an n -cycle. A finite chain C such that $C = \partial B$ is an n -boundary. Let $Z_n(M) = \text{Ker}(\partial)$ (resp. $B_n(M) = \text{Im}(\partial)$) denote a real vector space of cycles (resp. boundaries) of degree n . Since $\partial_{n+1}\partial_n = \partial^2 = 0$, we have $B_n(M) \subset Z_n(M)$. The quotient vector space

$$H_n^C(M) = \text{Ker}(\partial)/\text{Im}(\partial) = Z_n(M)/B_n(M)$$

is the n -homology group. The elements of $H_n^C(M)$ are equivalence sets of cycles. Two cycles C_1, C_2 belong to the same equivalence set, or are homologous (written $C_1 \sim C_2$), iff they differ by a boundary $C_1 - C_2 = \partial B$. The homology class of a finite chain $C \in C^n(M)$ is $[C] \in H_n^C(M)$.

The dimension of the n -cohomology (resp. n -homology) group equals the n th Betti number b^n (resp. b_n) of the manifold M . *Poincaré lemma* says that on an open set $U \in M$ diffeomorphic to \mathbb{R}^N , all closed forms (cycles) of degree $p \geq 1$ are exact (boundaries). That is, the Betti numbers satisfy $b^p = 0$ (resp. $b_p = 0$) for $p = 1, \dots, n$.

The *de Rham Theorem* states the following. The map $\Phi: H_n \times H^n \rightarrow \mathbb{R}$ given by $([C], [\omega]) \rightarrow \langle C, \omega \rangle$ for $C \in Z_n, \omega \in Z^n$ is a bilinear nondegenerate map which establishes the duality of the groups (vector spaces) H_n and H^n and the equality $b_n = b^n$.

3.6.3.7 Hodge Star Operator and Harmonic Forms

As the configuration manifold M is an oriented ND Riemannian manifold, we may select an orientation on all tangent spaces $T_m M$ and all cotangent spaces $T_m^* M$, with the local coordinates $x^i = (q^i, p_i)$ at a point $m \in M$, in a consistent manner. The simplest way to do that is to choose the Euclidean orthonormal basis $\partial_1, \dots, \partial_N$ of \mathbb{R}^N as being positive.

Since the manifold M carries a Riemannian structure $g = \langle \cdot, \cdot \rangle$, we have a scalar product on each $T_m^* M$. So, we can define (as above) the linear

Hodge star operator

$$* : \Lambda^p(T_m^* M) \rightarrow \Lambda^{N-p}(T_m^* M),$$

which is a base point preserving operator

$$* : \Omega^p(M) \rightarrow \Omega^{N-p}(M), \quad (\Omega^p(M) = \Gamma(\Lambda^p(M)))$$

(here $\Lambda^p(V)$ denotes the p -fold exterior product of any vector space V , $\Omega^p(M)$ is a space of all p -forms on M , and $\Gamma(E)$ denotes the space of sections of the vector bundle E). Also,

$$** = (-1)^{p(N-p)} : \Lambda^p(T_x^* M) \rightarrow \Lambda^p(T_x^* M).$$

As the metric on $T_m^* M$ is given by $g^{ij}(x) = (g_{ij}(x))^{-1}$, we have the *volume form* defined in local coordinates as

$$*(1) = \sqrt{\det(g_{ij})} dx^1 \wedge \dots \wedge dx^n, \quad \text{and} \quad \text{Vol}(M) = \int_M *(1).$$

For any to p -forms $\alpha, \beta \in \Omega^p(M)$ with compact support, we define the (bilinear and positive definite) L^2 -product as

$$(\alpha, \beta) = \int_M \langle \alpha, \beta \rangle * (1) = \int_M \alpha \wedge *\beta.$$

We can extend the product (\cdot, \cdot) to $L^2(\Omega^p(M))$; it remains bilinear and positive definite, because as usual, in the definition of L^2 , functions that differ only on a set of measure zero are identified.

Using the Hodge star operator $*$, we can introduce the *codifferential* operator δ , which is formally adjoint to the exterior derivative $d : \Omega^p(M) \rightarrow \Omega^{p+1}(M)$ on $\oplus_{p=0}^N \Omega^p(M)$ w.r.t. (\cdot, \cdot) . This means that for $\alpha \in \Omega^{p-1}(M), \beta \in \Omega^p(M)$

$$(d\alpha, \beta) = (\alpha, \delta\beta).$$

Therefore, we have $\delta : \Omega^p(M) \rightarrow \Omega^{p-1}(M)$ and

$$\delta = (-1)^{N(p+1)+1} * d *.$$

Now, the *Laplace–Beltrami operator* (or, *Hodge Laplacian*, see [Griffiths (1983b); Voisin (2002)] as well as section (3.13.5.3) below), Δ on $\Omega^p(M)$, is defined by relation similar to (3.44) above

$$\Delta = d\delta + \delta d : \Omega^p(M) \rightarrow \Omega^p(M) \quad (3.47)$$

and an exterior differential form $\alpha \in \Omega^p(M)$ is called *harmonic* if $\Delta\alpha = 0$.

Let M be a compact, oriented Riemannian manifold, E a vector bundle with a bundle metric $\langle \cdot, \cdot \rangle$ over M ,

$$D = d + A : \Omega^{p-1}(Ad_E) \rightarrow \Omega^p(Ad_E), \quad \text{with } A \in \Omega^1(Ad_E)$$

– a tensorial and \mathbb{R} –linear metric connection on E with curvature $F_D \in \Omega^2(Ad_E)$ (Here by $\Omega^p(Ad_E)$ we denote the space of those elements of $\Omega^p(End_E)$ for which the endomorphism of each fibre is skew symmetric; End_E denotes the space of linear endomorphisms of the fibers of E).

3.7 Lie Derivatives on Smooth Manifolds

Lie derivative is popularly called ‘*fisherman’s derivative*’. In continuum mechanics it is called *Liouville operator*. This is a central differential operator in modern differential geometry and its physical and control applications.

3.7.1 Lie Derivative Operating on Functions

To define how vector–fields operate on functions on an m –manifold M , we will use the *Lie derivative*. Let $f : M \rightarrow \mathbb{R}$ so $Tf : TM \rightarrow T\mathbb{R} = \mathbb{R} \times \mathbb{R}$. Following [Abraham *et al.* (1988)] we write Tf acting on a vector $v \in T_m M$ in the form

$$Tf \cdot v = (f(m), df(m) \cdot v).$$

This defines, for each point $m \in M$, the element $df(m) \in T_m^* M$. Thus df is a section of the cotangent bundle T^*M , i.e., a 1–form. The 1–form $df : M \rightarrow T^*M$ defined this way is called the *differential* of f . If f is C^k , then df is C^{k-1} .

If $\phi : U \subset M \rightarrow V \subset E$ is a local chart for M , then the local representative of $f \in C^k(M, \mathbb{R})$ is the map $f : V \rightarrow \mathbb{R}$ defined by $f = f \circ \phi^{-1}$. The local representative of Tf is the tangent map for local manifolds,

$$Tf(x, v) = (f(x), Df(x) \cdot v).$$

Thus the local representative of df is the derivative of the local representative of f . In particular, if (x^1, \dots, x^n) are local coordinates on M , then the local components of df are

$$(df)^i = \partial_{x^i} f.$$

The introduction of df leads to the following definition of the Lie derivative. The *directional* or *Lie derivative* $\mathcal{L}_X : C^k(M, \mathbb{R}) \rightarrow C^{k-1}(M, \mathbb{R})$ of a function $f \in C^k(M, \mathbb{R})$ along a vector-field X is defined by

$$\mathcal{L}_X f(m) = X[f](m) = df(m) \cdot X(m),$$

for any $m \in M$. Denote by $X[f] = df(X)$ the map $M \ni m \mapsto X[f](m) \in \mathbb{R}$. If f is F -valued, the same definition is used, but now $X[f]$ is F -valued.

If a local chart (U, ϕ) on an n -manifold M has local coordinates (x^1, \dots, x^n) , the local representative of $X[f]$ is given by the function

$$\mathcal{L}_X f = X[f] = X^i \partial_{x^i} f.$$

Evidently if f is C^k and X is C^{k-1} then $X[f]$ is C^{k-1} .

Let $\varphi : M \rightarrow N$ be a diffeomorphism. Then \mathcal{L}_X is natural with respect to push-forward by φ . That is, for each $f \in C^k(M, \mathbb{R})$,

$$\mathcal{L}_{\varphi_* X}(\varphi_* f) = \varphi_* \mathcal{L}_X f,$$

i.e., the following diagram commutes:

$$\begin{array}{ccc} C^k(M, \mathbb{R}) & \xrightarrow{\varphi_*} & C^k(N, \mathbb{R}) \\ \mathcal{L}_X \downarrow & & \downarrow \mathcal{L}_{\varphi_* X} \\ C^k(M, \mathbb{R}) & \xrightarrow{\varphi_*} & C^k(N, \mathbb{R}) \end{array}$$

Also, \mathcal{L}_X is natural with respect to restrictions. That is, for U open in M and $f \in C^k(M, \mathbb{R})$,

$$\mathcal{L}_{X|U}(f|U) = (\mathcal{L}_X f)|U,$$

where $|U : C^k(M, \mathbb{R}) \rightarrow C^k(U, \mathbb{R})$ denotes restriction to U , i.e., the following diagram commutes:

$$\begin{array}{ccc} C^k(M, \mathbb{R}) & \xrightarrow{|U} & C^k(U, \mathbb{R}) \\ \mathcal{L}_X \downarrow & & \downarrow \mathcal{L}_{X|U} \\ C^k(M, \mathbb{R}) & \xrightarrow{|U} & C^k(U, \mathbb{R}) \end{array}$$

Since $\varphi^* = (\varphi^{-1})_*$ the Lie derivative is also natural with respect to pull-back by φ . This has a generalization to φ -related vector-fields as

follows: Let $\varphi : M \rightarrow N$ be a C^k -map, $X \in \mathcal{X}^{k-1}(M)$ and $Y \in \mathcal{X}^{k-1}(N)$, $k \geq 1$. If $X \sim_\varphi Y$, then

$$\mathcal{L}_X(\varphi^* f) = \varphi^* \mathcal{L}_Y f$$

for all $f \in C^k(N, \mathbb{R})$, i.e., the following diagram commutes:

$$\begin{array}{ccc} C^k(N, \mathbb{R}) & \xrightarrow{\varphi^*} & C^k(M, \mathbb{R}) \\ \mathcal{L}_Y \downarrow & & \downarrow \mathcal{L}_X \\ C^k(N, \mathbb{R}) & \xrightarrow{\varphi^*} & C^k(M, \mathbb{R}) \end{array}$$

The Lie derivative map $\mathcal{L}_X : C^k(M, \mathbb{R}) \rightarrow C^{k-1}(M, \mathbb{R})$ is a *derivation*, i.e., for two functions $f, g \in C^k(M, \mathbb{R})$ the *Leibniz rule* is satisfied

$$\mathcal{L}_X(fg) = g\mathcal{L}_X f + f\mathcal{L}_X g;$$

Also, Lie derivative of a constant function is zero, $\mathcal{L}_X(\text{const}) = 0$.

The connection between the Lie derivative $\mathcal{L}_X f$ of a function $f \in C^k(M, \mathbb{R})$ and the flow F_t of a vector-field $X \in \mathcal{X}^{k-1}(M)$ is given as:

$$\frac{d}{dt} (F_t^* f) = F_t^* (\mathcal{L}_X f).$$

3.7.2 Lie Derivative of Vector Fields

If $X, Y \in \mathcal{X}^k(M)$, $k \geq 1$ are two vector-fields on M , then

$$[\mathcal{L}_X, \mathcal{L}_Y] = \mathcal{L}_X \circ \mathcal{L}_Y - \mathcal{L}_Y \circ \mathcal{L}_X$$

is a derivation map from $C^{k+1}(M, \mathbb{R})$ to $C^{k-1}(M, \mathbb{R})$. Then there is a unique vector-field, $[X, Y] \in \mathcal{X}^k(M)$ of X and Y such that $\mathcal{L}_{[X, Y]} = [\mathcal{L}_X, \mathcal{L}_Y]$ and $[X, Y](f) = X(Y(f)) - Y(X(f))$ holds for all functions $f \in C^k(M, \mathbb{R})$. This vector-field is also denoted $\mathcal{L}_X Y$ and is called the Lie derivative of Y with respect to X , or the *Lie bracket* of X and Y . In a local chart (U, ϕ) at a point $m \in M$ with coordinates (x^1, \dots, x^n) , for $X|_U = X^i \partial_{x^i}$ and $Y|_U = Y^j \partial_{x^j}$ we have

$$[X^i \partial_{x^i}, Y^j \partial_{x^j}] = (X^i (\partial_{x^i} Y^j) - Y^j (\partial_{x^j} X^i)) \partial_{x^j},$$

since second partials commute. If, also X has flow F_t , then [Abraham *et al.* (1988)]

$$\frac{d}{dt} (F_t^* Y) = F_t^* (\mathcal{L}_X Y).$$

In particular, if $t = 0$, this formula becomes

$$\frac{d}{dt} \big|_{t=0} (F_t^* Y) = \mathcal{L}_X Y.$$

Then the unique C^{k-1} vector-field $\mathcal{L}_X Y = [X, Y]$ on M defined by

$$[X, Y] = \frac{d}{dt} \big|_{t=0} (F_t^* Y),$$

is called the Lie derivative of Y with respect to X , or the Lie bracket of X and Y , and can be interpreted as the leading order term that results from the sequence of flows

$$F_t^{-Y} \circ F_t^{-X} \circ F_t^Y \circ F_t^{-X}(m) = \epsilon^2 [X, Y](m) + \mathcal{O}(\epsilon^3), \quad (3.48)$$

for some real $\epsilon > 0$. Therefore a Lie bracket can be interpreted as a ‘new direction’ in which the system can flow, by executing the sequence of flows (3.48).

Lie bracket satisfies the following property:

$$[X, Y][f] = X[Y[f]] - Y[X[f]],$$

for all $f \in C^{k+1}(U, \mathbb{R})$, where U is open in M .

An important relationship between flows of vector-fields is given by the *Campbell–Baker–Hausdorff formula*:

$$F_t^Y \circ F_t^X = F_t^{X+Y+\frac{1}{2}[X,Y]+\frac{1}{12}([X,[X,Y]]-[Y,[X,Y]])+\dots} \quad (3.49)$$

Essentially, if given the composition of multiple flows along multiple vector-fields, this formula gives the one flow along one vector-field which results in the same net flow. One way to prove the Campbell–Baker–Hausdorff formula (3.49) is to expand the product of two formal exponentials and equate terms in the resulting formal power series.

Lie bracket is the \mathbb{R} -bilinear map $[\cdot, \cdot] : \mathcal{X}^k(M) \times \mathcal{X}^k(M) \rightarrow \mathcal{X}^k(M)$ with the following properties:

- (1) $[X, Y] = -[Y, X]$, i.e., $\mathcal{L}_X Y = -\mathcal{L}_Y X$ for all $X, Y \in \mathcal{X}^k(M)$ – skew-symmetry;
- (2) $[X, X] = 0$ for all $X \in \mathcal{X}^k(M)$;

- (3) $[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0$ for all $X, Y, Z \in \mathcal{X}^k(M)$ – the Jacobi identity;
- (4) $[fX, Y] = f[X, Y] - (Yf)X$, i.e., $\mathcal{L}_{fX}(Y) = f(\mathcal{L}_X Y) - (\mathcal{L}_Y f)X$ for all $X, Y \in \mathcal{X}^k(M)$ and $f \in C^k(M, \mathbb{R})$;
- (5) $[X, fY] = f[X, Y] + (Xf)Y$, i.e., $\mathcal{L}_X(fY) = f(\mathcal{L}_X Y) + (\mathcal{L}_X f)Y$ for all $X, Y \in \mathcal{X}^k(M)$ and $f \in C^k(M, \mathbb{R})$;
- (6) $[\mathcal{L}_X, \mathcal{L}_Y] = \mathcal{L}_{[X, Y]}$ for all $X, Y \in \mathcal{X}^k(M)$.

The pair $(\mathcal{X}^k(M), [\cdot, \cdot])$ is the prototype of a *Lie algebra* [Kolar *et al.* (1993)]. In more general case of a *general linear Lie algebra* $\mathfrak{gl}(n)$, which is the Lie algebra associated to the Lie group $GL(n)$, Lie bracket is given by a *matrix commutator*

$$[A, B] = AB - BA,$$

for any two matrices $A, B \in \mathfrak{gl}(n)$.

Let $\varphi : M \rightarrow N$ be a diffeomorphism. Then $\mathcal{L}_X : \mathcal{X}^k(M) \rightarrow \mathcal{X}^k(M)$ is natural with respect to push-forward by φ . That is, for each $f \in C^k(M, \mathbb{R})$,

$$\mathcal{L}_{\varphi_* X}(\varphi_* Y) = \varphi_* \mathcal{L}_X Y,$$

i.e., the following diagram commutes:

$$\begin{array}{ccc} \mathcal{X}^k(M) & \xrightarrow{\varphi_*} & \mathcal{X}^k(N) \\ \mathcal{L}_X \downarrow & & \downarrow \mathcal{L}_{\varphi_* X} \\ \mathcal{X}^k(M) & \xrightarrow{\varphi_*} & \mathcal{X}^k(N) \end{array}$$

Also, \mathcal{L}_X is natural with respect to restrictions. That is, for U open in M and $f \in C^k(M, \mathbb{R})$,

$$[X|U, Y|U] = [X, Y]|U,$$

where $U : C^k(M, \mathbb{R}) \rightarrow C^k(U, \mathbb{R})$ denotes restriction to U , i.e., the following diagram commutes [Abraham *et al.* (1988)]:

$$\begin{array}{ccc} \mathcal{X}^k(M) & \xrightarrow{|U} & \mathcal{X}^k(U) \\ \mathcal{L}_X \downarrow & & \downarrow \mathcal{L}_{X|U} \\ \mathcal{X}^k(M) & \xrightarrow{|U} & \mathcal{X}^k(U) \end{array}$$

If a local chart (U, ϕ) on an n -manifold M has local coordinates (x^1, \dots, x^n) , then the local components of a Lie bracket are

$$[X, Y]^j = X^i \partial_{x^i} Y^j - Y^i \partial_{x^i} X^j,$$

that is, $[X, Y] = (X \cdot \nabla)Y - (Y \cdot \nabla)X$.

Let $\varphi : M \rightarrow N$ be a C^k -map, $X \in \mathcal{X}^{k-1}(M)$ and $Y \in \mathcal{X}^{k-1}(N)$, $k \geq 1$. Then $X \sim_\varphi Y$, iff

$$(Y[f]) \circ \varphi = X[f \circ \varphi]$$

for all $f \in C^k(V, \mathbb{R})$, where V is open in N .

For every $X \in \mathcal{X}^k(M)$, the operator \mathcal{L}_X is a derivation on $(C^k(M, \mathbb{R}), \mathcal{X}^k(M))$, i.e., \mathcal{L}_X is \mathbb{R} -linear.

For any two vector-fields $X \in \mathcal{X}^k(M)$ and $Y \in \mathcal{X}^k(N)$, $k \geq 1$ with flows F_t and G_t , respectively, if $[X, Y] = 0$ then $F_t^* Y = Y$ and $G_t^* X = X$.

3.7.3 Time Derivative of the Evolution Operator

Recall that the time-dependent flow or evolution operator $F_{t,s}$ of a vector-field $X \in \mathcal{X}^k(M)$ is defined by the requirement that $t \mapsto F_{t,s}(m)$ be the integral curve of X starting at a point $m \in M$ at time $t = s$, i.e.,

$$\frac{d}{dt} F_{t,s}(m) = X(t, F_{t,s}(m)) \quad \text{and} \quad F_{t,t}(m) = m.$$

By uniqueness of integral curves we have $F_{t,s} \circ F_{s,r} = F_{t,r}$ (replacing the flow property $F_{t+s} = F_t + F_s$) and $F_{t,t} = \text{identity}$.

Let $X_t \in \mathcal{X}^k(M)$, $k \geq 1$ for each t and suppose $X(t, m)$ is continuous in $(t, m) \in \mathbb{R} \times M$. Then $F_{t,s}$ is of class C^k and for $f \in C^{k+1}(M, \mathbb{R})$ [Abraham *et al.* (1988)], and $Y \in \mathcal{X}^k(M)$, we have

- (1) $\frac{d}{dt} F_{t,s}^* f = F_{t,s}^* (\mathcal{L}_{X_t} f)$, and
- (2) $\frac{d}{dt} F_{t,s}^* f = F_{t,s}^* ([X_t, Y]) = F_{t,s}^* (\mathcal{L}_{X_t} Y)$.

From the above Theorem, the following identity holds:

$$\frac{d}{dt} F_{t,s}^* f = -X_t [F_{t,s}^* f].$$

3.7.4 Lie Derivative of Differential Forms

Since $\mathcal{F} : M \implies \Lambda^k T^* M$ is a vector bundle functor on \mathcal{M} , the Lie derivative (4.3.2) of a k -form $\alpha \in \Omega^k(M)$ along a vector-field $X \in \mathcal{X}^k(M)$ is

defined by

$$\mathcal{L}_X \alpha = \frac{d}{dt} \Big|_{t=0} F_t^* \alpha.$$

It has the following properties:

- (1) $\mathcal{L}_X(\alpha \wedge \beta) = \mathcal{L}_X \alpha \wedge \beta + \alpha \wedge \mathcal{L}_X \beta$, so \mathcal{L}_X is a derivation.
- (2) $[\mathcal{L}_X, \mathcal{L}_Y] \alpha = \mathcal{L}_{[X, Y]} \alpha$.
- (3) $\frac{d}{dt} F_t^* \alpha = F_t^* \mathcal{L}_X \alpha = \mathcal{L}_X (F_t^* \alpha)$.

Formula (3) holds also for time-dependent vector-fields in the sense that $\frac{d}{dt} F_{t,s}^* \alpha = F_{t,s}^* \mathcal{L}_X \alpha = \mathcal{L}_X (F_{t,s}^* \alpha)$ and in the expression $\mathcal{L}_X \alpha$ the vector-field X is evaluated at time t .

The famous *Cartan magic formula* (see [Marsden and Ratiu (1999)]) states: the Lie derivative of a k -form $\alpha \in \Omega^k(M)$ along a vector-field $X \in \mathcal{X}^k(M)$ on a smooth manifold M is defined as

$$\mathcal{L}_X \alpha = di_X \alpha + i_X d\alpha = d(X \lrcorner \alpha) + X \lrcorner d\alpha.$$

Also, the following identities hold [Marsden and Ratiu (1999); Kolar *et al.* (1993)]:

- (1) $\mathcal{L}_f X \alpha = f \mathcal{L}_X \alpha + df \wedge i_X \alpha$.
- (2) $\mathcal{L}_{[X, Y]} \alpha = \mathcal{L}_X \mathcal{L}_Y \alpha - \mathcal{L}_Y \mathcal{L}_X \alpha$.
- (3) $i_{[X, Y]} \alpha = \mathcal{L}_X i_Y \alpha - i_Y \mathcal{L}_X \alpha$.
- (4) $\mathcal{L}_X d\alpha = d\mathcal{L}_X \alpha$, i.e., $[\mathcal{L}_X, d] = 0$.
- (5) $\mathcal{L}_X i_X \alpha = i_X \mathcal{L}_X \alpha$, i.e., $[\mathcal{L}_X, i_X] = 0$.
- (6) $\mathcal{L}_X(\alpha \wedge \beta) = \mathcal{L}_X \alpha \wedge \beta + \alpha \wedge \mathcal{L}_X \beta$.

3.7.5 Lie Derivative of Various Tensor Fields

In this section, we use local coordinates x^i ($i = 1, \dots, n$) on a biodynamical n -manifold M , to calculate the Lie derivative \mathcal{L}_{X^i} with respect to a generic vector-field X^i . (As always, $\partial_{x^i} \equiv \frac{\partial}{\partial x^i}$).

Lie Derivative of a Scalar Field

Given the scalar field ϕ , its Lie derivative $\mathcal{L}_{X^i} \phi$ is given as

$$\mathcal{L}_{X^i} \phi = X^i \partial_{x^i} \phi = X^1 \partial_{x^1} \phi + X^2 \partial_{x^2} \phi + \dots + X^n \partial_{x^n} \phi.$$

Lie Derivative of Vector and Covector-Fields

Given a contravariant vector-field V^i , its Lie derivative $\mathcal{L}_{X^i}V^i$ is given as

$$\mathcal{L}_{X^i}V^i = X^k\partial_{x^k}V^i - V^k\partial_{x^k}X^i \equiv [X^i, V^i] - \text{the Lie bracket.}$$

Given a covariant vector-field (i.e., a one-form) ω_i , its Lie derivative $\mathcal{L}_{X^i}\omega_i$ is given as

$$\mathcal{L}_{X^i}\omega_i = X^k\partial_{x^k}\omega_i + \omega_k\partial_{x^i}X^k.$$

Lie Derivative of a Second-Order Tensor-Field

Given a $(2, 0)$ tensor-field S^{ij} , its Lie derivative $\mathcal{L}_{X^i}S^{ij}$ is given as

$$\mathcal{L}_{X^i}S^{ij} = X^i\partial_{x^i}S^{ij} - S^{ij}\partial_{x^i}X^i - S^{ii}\partial_{x^i}X^j.$$

Given a $(1, 1)$ tensor-field S_j^i , its Lie derivative $\mathcal{L}_{X^i}S_j^i$ is given as

$$\mathcal{L}_{X^i}S_j^i = X^i\partial_{x^i}S_j^i - S_j^i\partial_{x^i}X^i + S_i^i\partial_{x^j}X^i.$$

Given a $(0, 2)$ tensor-field S_{ij} , its Lie derivative $\mathcal{L}_{X^i}S_{ij}$ is given as

$$\mathcal{L}_{X^i}S_{ij} = X^i\partial_{x^i}S_{ij} + S_{ij}\partial_{x^i}X^i + S_{ii}\partial_{x^j}X^i.$$

Lie Derivative of a Third-Order Tensor-Field

Given a $(3, 0)$ tensor-field T^{ijk} , its Lie derivative $\mathcal{L}_{X^i}T^{ijk}$ is given as

$$\mathcal{L}_{X^i}T^{ijk} = X^i\partial_{x^i}T^{ijk} - T^{ijk}\partial_{x^i}X^i - T^{iik}\partial_{x^i}X^j - T^{ijji}\partial_{x^i}X^k.$$

Given a $(2, 1)$ tensor-field T_k^{ij} , its Lie derivative $\mathcal{L}_{X^i}T_k^{ij}$ is given as

$$\mathcal{L}_{X^i}T_k^{ij} = X^i\partial_{x^i}T_k^{ij} - T_k^{ij}\partial_{x^i}X^i + T_i^{ij}\partial_{x^k}X^i - T_k^{ii}\partial_{x^i}X^j.$$

Given a $(1, 2)$ tensor-field T_{jk}^i , its Lie derivative $\mathcal{L}_{X^i}T_{jk}^i$ is given as

$$\mathcal{L}_{X^i}T_{jk}^i = X^i\partial_{x^i}T_{jk}^i - T_{jk}^i\partial_{x^i}X^i + T_{ik}^i\partial_{x^j}X^i + T_{ji}^i\partial_{x^k}X^i.$$

Given a $(0, 3)$ tensor-field T_{ijk} , its Lie derivative $\mathcal{L}_{X^i}T_{ijk}$ is given as

$$\mathcal{L}_{X^i}T_{ijk} = X^i\partial_{x^i}T_{ijk} + T_{ijk}\partial_{x^i}X^i + T_{iik}\partial_{x^j}X^i + T_{ijji}\partial_{x^k}X^i.$$

Lie Derivative of a Fourth-Order Tensor-Field

Given a $(4, 0)$ tensor-field R^{ijkl} , its Lie derivative $\mathcal{L}_{X^i} R^{ijkl}$ is given as

$$\mathcal{L}_{X^i} R^{ijkl} = X^i \partial_{x^i} R^{ijkl} - R^{ijkl} \partial_{x^i} X^i - R^{ikl} \partial_{x^i} X^j - R^{ijl} \partial_{x^i} X^k - R^{ijk} \partial_{x^i} X^l.$$

Given a $(3, 1)$ tensor-field R_l^{ijk} , its Lie derivative $\mathcal{L}_{X^i} R_l^{ijk}$ is given as

$$\mathcal{L}_{X^i} R_l^{ijk} = X^i \partial_{x^i} R_l^{ijk} - R_l^{ijk} \partial_{x^i} X^i + R_l^{ijk} \partial_{x^i} X^j - R_l^{iik} \partial_{x^i} X^j - R_l^{iji} \partial_{x^i} X^k.$$

Given a $(2, 2)$ tensor-field R_{kl}^{ij} , its Lie derivative $\mathcal{L}_{X^i} R_{kl}^{ij}$ is given as

$$\mathcal{L}_{X^i} R_{kl}^{ij} = X^i \partial_{x^i} R_{kl}^{ij} - R_{kl}^{ij} \partial_{x^i} X^i + R_{il}^{ij} \partial_{x^k} X^i + R_{ki}^{ij} \partial_{x^l} X^i - R_{kl}^{ii} \partial_{x^i} X^j.$$

Given a $(1, 3)$ tensor-field R_{jkl}^i , its Lie derivative $\mathcal{L}_{X^i} R_{jkl}^i$ is given as

$$\mathcal{L}_{X^i} R_{jkl}^i = X^i \partial_{x^i} R_{jkl}^i - R_{jkl}^i \partial_{x^i} X^i + R_{ikl}^i \partial_{x^j} X^i + R_{jil}^i \partial_{x^k} X^i + R_{jki}^i \partial_{x^l} X^i.$$

Given a $(0, 4)$ tensor-field R_{ijkl} , its Lie derivative $\mathcal{L}_{X^i} R_{ijkl}$ is given as

$$\mathcal{L}_{X^i} R_{ijkl} = X^i \partial_{x^i} R_{ijkl} + R_{ijkl} \partial_{x^i} X^i + R_{iikl} \partial_{x^j} X^i + R_{ijil} \partial_{x^k} X^i + R_{ijki} \partial_{x^l} X^i.$$

Finally, recall that a *spinor* is a two-component complex column vector. Physically, spinors can describe both bosons and fermions, while tensors can describe only bosons. The Lie derivative of a spinor ϕ is defined by

$$\mathcal{L}_X \phi(x) = \lim_{t \rightarrow 0} \frac{\bar{\phi}_t(x) - \phi(x)}{t},$$

where $\bar{\phi}_t$ is the image of ϕ by a one-parameter group of isometries with X its generator. For a vector-field X^a and a covariant derivative ∇_a , the Lie derivative of ϕ is given explicitly by

$$\mathcal{L}_X \phi = X^a \nabla_a \phi - \frac{1}{8} (\nabla_a X_b - \nabla_b X_a) \gamma^a \gamma^b \phi,$$

where γ^a and γ^b are *Dirac matrices* (see, e.g., [Choquet-Bruhat and DeWitt-Morete (2000)]).

3.7.6 Application: Lie-Derivative Neurodynamics

A *Lie-derivative neuro-classifier* is a self-organized, associative-memory machine, represented by oscillatory (excitatory/inhibitory) tensor-field-system

(x, v, ω) on the Banach manifold M . It consists of continual neural activation (x, y) -dynamics and self-organizing synaptic learning ω -dynamics.

The continual activation (x, y) –dynamics, is defined as a system of two coupled, first–order oscillator tensor–fields, dual to each other, in a local Banach chart U_α , $(\alpha = 1, \dots, n)$ on M :

1) an excitatory neural vector–field $x^i = x^i(t) : M \rightarrow TM$, representing a cross–section of the tangent bundle TM ; and

2) an inhibitory neural one–form $y_i = y_i(t) : M \rightarrow T^*M$, representing a cross–section of the cotangent bundle T^*M .

The self–organized learning ω –dynamics is performed on a second–order symmetrical synaptic tensor–field $\omega = \omega(t)$, given by its covariant components $\omega_{ij} = \omega_{ij}(t)$ and its contravariant components $\omega^{ij} = \omega^{ij}(t)$, where $i, j = 1, \dots, n$.

Starting with the Lyapunov–stable, negative scalar neural action potential:

$$U = -\frac{1}{2}(\omega_{ij}x^ix^j + \omega^{ij}y_iy_j), \quad (i, j = 1, \dots, n),$$

the (x, y) –dynamics is given in two versions, which are compared and contrasted:

(1) the **Lie–linear** neurodynamics with first–order Lie derivatives

$$\dot{x}^i = J^i + L_X U, \quad \dot{y}_i = J_i + L_Y U,$$

and

(2) the **Lie–quadratic** neurodynamics with both first and second–order Lie derivatives

$$\dot{x}^i = J^i + L_X U + L_X L_X U, \quad \dot{y}_i = J_i + L_Y U + L_Y L_Y U,$$

where $X = S_i(x^i)$, $Y = S_i(y_i)$, S_i represent sigmoid activation functions, while $L_X L_X, L_Y L_Y : F(M) \rightarrow F(M)$ denote the *second–order (iterated) Lie derivatives*.

Self–organized learning ω –dynamics is presented in the form of differential Hebbian learning scheme in both covariant and contravariant forms

$$\begin{aligned} \dot{\omega}_{ij} &= -\omega_{ij} + S_i(x^i)S_j(y_j) + \dot{S}_i(x^i)\dot{S}_j(y_j), & \text{and} \\ \dot{\omega}^{ij} &= -\omega^{ij} + S_i(x^i)S_j(y_j) + \dot{S}_i(x^i)\dot{S}_j(y_j), & (i, j = 1, \dots, n), \end{aligned}$$

respectively.

3.7.7 Lie Algebras

Recall from Introduction that an *algebra* A is a vector space with a product. The product must have the property that

$$a(uv) = (au)v = u(av),$$

for every $a \in \mathbb{R}$ and $u, v \in A$. A map $\phi : A \rightarrow A'$ between algebras is called an *algebra homomorphism* if $\phi(u \cdot v) = \phi(u) \cdot \phi(v)$. A vector subspace \mathfrak{I} of an algebra A is called a *left ideal* (resp. *right ideal*) if it is closed under algebra multiplication and if $u \in A$ and $i \in \mathfrak{I}$ implies that $ui \in \mathfrak{I}$ (resp. $iu \in \mathfrak{I}$). A subspace \mathfrak{I} is said to be a *two-sided ideal* if it is both a left and right ideal. An ideal may not be an algebra itself, but the quotient of an algebra by a two-sided ideal inherits an algebra structure from A .

A *Lie algebra* is an algebra A where the multiplication, i.e., the *Lie bracket* $(u, v) \mapsto [u, v]$, has the following properties:

LA 1. $[u, u] = 0$ for every $u \in A$, and

LA 2. $[u, [v, w]] + [w, [u, v]] + [v, w, u] = 0$ for all $u, v, w \in A$.

The condition LA 2 is usually called *Jacobi identity*. A subspace $E \subset A$ of a Lie algebra is called a *Lie subalgebra* if $[u, v] \in E$ for every $u, v \in E$. A map $\phi : A \rightarrow A'$ between Lie algebras is called a *Lie algebra homomorphism* if $\phi([u, v]) = [\phi(u), \phi(v)]$ for each $u, v \in A$.

All Lie algebras (over a given field \mathbb{K}) and all smooth homomorphisms between them form the category \mathcal{LAL} , which is itself a complete subcategory of the category \mathcal{AL} of all algebras and their homomorphisms.

3.8 Lie Groups and Associated Lie Algebras

In the middle of the 19th Century S. Lie made a far reaching discovery that techniques designed to solve particular unrelated types of ODEs, such as separable, homogeneous and exact equations, were in fact all special cases of a general form of integration procedure based on the invariance of the differential equation under a continuous group of symmetries. Roughly speaking a symmetry group of a system of differential equations is a group that transforms solutions of the system to other solutions. Once the symmetry group has been identified a number of techniques to solve and classify these differential equations becomes possible. In the classical framework of Lie, these groups were local groups and arose locally as groups of transformations on some Euclidean space. The passage from the local Lie group to the present day definition using manifolds was accomplished by E. Cartan

at the end of the 19th Century, whose work is a striking synthesis of Lie theory, classical geometry, differential geometry and topology.

These continuous groups, which originally appeared as symmetry groups of differential equations, have over the years had a profound impact on diverse areas such as algebraic topology, differential geometry, numerical analysis, control theory, classical mechanics, quantum mechanics etc. They are now universally known as Lie groups.

3.8.1 Definition of a Lie Group

A *Lie group* is a smooth (Banach) manifold M that has at the same time a group G -structure consistent with its manifold M -structure in the sense that *group multiplication*

$$\mu : G \times G \rightarrow G, \quad (g, h) \mapsto gh \quad (3.50)$$

and the *group inversion*

$$\nu : G \rightarrow G, \quad g \mapsto g^{-1} \quad (3.51)$$

are C^k -maps [Chevalley (1955); Abraham *et al.* (1988); Marsden and Ratiu (1999); Puta (1993)]. A point $e \in G$ is called the *group identity element*.

For example, any n D Banach vector space V is an Abelian Lie group with group operations $\mu : V \times V \rightarrow V$, $\mu(x, y) = x + y$, and $\nu : V \rightarrow V$, $\nu(x) = -x$. The identity is just the zero vector. We call such a Lie group a *vector group*.

Let G and H be two Lie groups. A map $G \rightarrow H$ is said to be a *morphism* of Lie groups (or their *smooth homomorphism*) if it is their homomorphism as abstract groups and their smooth map as manifolds [Postnikov (1986)].

All Lie groups and all their morphisms form the category \mathcal{LG} (more precisely, there is a countable family of categories \mathcal{LG} depending on C^k -smoothness of the corresponding manifolds).

Similarly, a group G which is at the same time a topological space is said to be a *topological group* if maps (3.50–3.51) are continuous, i.e., C^0 -maps for it. The homomorphism $G \rightarrow H$ of topological groups is said to be continuous if it is a continuous map. Topological groups and their continuous homomorphisms form the category \mathcal{TG} .

A topological group (as well as a smooth manifold) is not necessarily Hausdorff. A topological group G is Hausdorff iff its identity is closed. As a corollary we have that every Lie group is a Hausdorff topological group (see [Postnikov (1986)]).

For every g in a Lie group G , the two maps,

$$\begin{aligned} L_g : G &\rightarrow G, & h &\mapsto gh, & \text{and} \\ R_h : G &\rightarrow G, & g &\mapsto gh, \end{aligned}$$

are called *left* and *right translation* maps. Since $L_g \circ L_h = L_{gh}$, and $R_g \circ R_h = R_{gh}$, it follows that $(L_g)^{-1} = L_{g^{-1}}$ and $(R_g)^{-1} = R_{g^{-1}}$, so both L_g and R_g are diffeomorphisms. Moreover $L_g \circ R_h = R_h \circ L_g$, i.e., left and right translation commute.

A vector-field X on G is called *left-invariant vector-field* if for every $g \in G$, $L_g^*X = X$, that is, if $(T_h L_g)X(h) = X(gh)$ for all $h \in G$, i.e., the following diagram commutes:

$$\begin{array}{ccc} TG & \xrightarrow{TL_g} & TG \\ \uparrow X & & \uparrow X \\ G & \xrightarrow{L_g} & G \end{array}$$

The correspondences $G \rightarrow TG$ and $L_g \rightarrow TL_g$ obviously define a functor $\mathcal{F} : \mathcal{LG} \Rightarrow \mathcal{LG}$ from the category G of Lie groups to itself. \mathcal{F} is a special case of the vector bundle functor (see (4.3.2) below).

Let $\mathcal{X}_L(G)$ denote the set of left-invariant vector-fields on G ; it is a Lie subalgebra of $\mathcal{X}(G)$, the set of all vector-fields on G , since $L_g^*[X, Y] = [L_g^*X, L_g^*Y] = [X, Y]$, so the Lie bracket $[X, Y] \in \mathcal{X}_L(G)$.

Let e be the identity element of G . Then for each ξ on the tangent space $T_e G$ we define a vector-field X_ξ on G by

$$X_\xi(g) = T_e L_g(\xi).$$

$\mathcal{X}_L(G)$ and $T_e G$ are isomorphic as vector spaces. Define the Lie bracket on $T_e G$ by

$$[\xi, \eta] = [X_\xi, X_\eta](e),$$

for all $\xi, \eta \in T_e G$. This makes $T_e G$ into a Lie algebra. Also, by construction, we have

$$[X_\xi, X_\eta] = X_{[\xi, \eta]},$$

this defines a bracket in $T_e G$ via *left extension*. The vector space $T_e G$ with the above algebra structure is called the Lie algebra of the Lie group G and is denoted \mathfrak{g} .

For example, let V be a nD vector space. Then $T_e V \simeq V$ and the left-invariant vector-field defined by $\xi \in T_e V$ is the constant vector-field $X_\xi(\eta) = \xi$, for all $\eta \in V$. The Lie algebra of V is V itself.

Since any two elements of an Abelian Lie group G commute, it follows that all adjoint operators Ad_g , $g \in G$, equal the identity. Therefore, the Lie algebra \mathfrak{g} is Abelian; that is, $[\xi, \eta] = 0$ for all $\xi, \eta \in \mathfrak{g}$ [Marsden and Ratiu (1999)].

Recall (3.7.7) that Lie algebras and their smooth homomorphisms form the category \mathcal{LAL} . We can now introduce the fundamental *Lie functor*, $\mathcal{F} : \mathcal{LG} \Rightarrow \mathcal{LAL}$, from the category of Lie groups to the category of Lie algebras [Postnikov (1986)].

Let X_ξ be a left-invariant vector-field on G corresponding to ξ in \mathfrak{g} . Then there is a unique integral curve $\gamma_\xi : \mathbb{R} \rightarrow G$ of X_ξ starting at e , i.e.,

$$\dot{\gamma}_\xi(t) = X_\xi(\gamma_\xi(t)), \quad \gamma_\xi(0) = e.$$

$\gamma_\xi(t)$ is a smooth *one-parameter subgroup* of G , i.e.,

$$\gamma_\xi(t+s) = \gamma_\xi(t) \cdot \gamma_\xi(s),$$

since, as functions of t both sides equal $\gamma_\xi(s)$ at $t = 0$ and both satisfy differential equation

$$\dot{\gamma}(t) = X_\xi(\gamma_\xi(t))$$

by left invariance of X_ξ , so they are equal. Left invariance can be also used to show that $\gamma_\xi(t)$ is defined for all $t \in \mathbb{R}$. Moreover, if $\phi : \mathbb{R} \rightarrow G$ is a one-parameter subgroup of G , i.e., a *smooth homomorphism* of the additive group \mathbb{R} into G , then $\phi = \gamma_\xi$ with $\xi = \dot{\phi}(0)$, since taking derivative at $s = 0$ in the relation

$$\phi(t+s) = \phi(t) \cdot \phi(s) \quad \text{gives} \quad \dot{\phi}(t) = X_{\dot{\phi}(0)}(\phi(t)),$$

so $\phi = \gamma_\xi$ since both equal e at $t = 0$. Therefore, all one-parameter subgroups of G are of the form $\gamma_\xi(t)$ for some $\xi \in \mathfrak{g}$.

The map $\exp : \mathfrak{g} \rightarrow G$, given by

$$\exp(\xi) = \gamma_\xi(1), \quad \exp(0) = e, \tag{3.52}$$

is called the *exponential map* of the Lie algebra \mathfrak{g} of G into G . \exp is a C^k -map, similar to the projection π of tangent and cotangent bundles; \exp is locally a diffeomorphism from a neighborhood of zero in \mathfrak{g} onto a

neighborhood of e in G ; if $f : G \rightarrow H$ is a smooth homomorphism of Lie groups, then

$$f \circ \exp_G = \exp_H \circ T_e f.$$

Also, in this case (see [Chevalley (1955); Marsden and Ratiu (1999); Postnikov (1986)])

$$\exp(s\xi) = \gamma_\xi(s).$$

Indeed, for fixed $s \in \mathbb{R}$, the curve $t \mapsto \gamma_\xi(ts)$, which at $t = 0$ passes through e , satisfies the differential equation

$$\frac{d}{dt} \gamma_\xi(ts) = sX_\xi(\gamma_\xi(ts)) = X_{s\xi}(\gamma_\xi(ts)).$$

Since $\gamma_{s\xi}(t)$ satisfies the same differential equation and passes through e at $t = 0$, it follows that $\gamma_{s\xi}(t) = \gamma_\xi(st)$. Putting $t = 1$ induces $\exp(s\xi) = \gamma_\xi(s)$ [Marsden and Ratiu (1999)].

Hence \exp maps the line $s\xi$ in \mathfrak{g} onto the one-parameter subgroup $\gamma_\xi(s)$ of G , which is tangent to ξ at e . It follows from left invariance that the flow F_t^ξ of X satisfies $F_t^\xi(g) = g \exp(s\xi)$.

Globally, the exponential map \exp , as given by (3.52), is a natural operation, i.e., for any morphism $\varphi : G \rightarrow H$ of Lie groups G and H and a Lie functor \mathcal{F} , the following diagram commutes [Postnikov (1986)]:

$$\begin{array}{ccc} \mathcal{F}(G) & \xrightarrow{\mathcal{F}(\varphi)} & \mathcal{F}(H) \\ \exp \downarrow & & \downarrow \exp \\ G & \xrightarrow{\varphi} & H \end{array}$$

Let G_1 and G_2 be Lie groups with Lie algebras \mathfrak{g}_1 and \mathfrak{g}_2 . Then $G_1 \times G_2$ is a Lie group with Lie algebra $\mathfrak{g}_1 \times \mathfrak{g}_2$, and the exponential map is given by [Marsden and Ratiu (1999)].

$$\exp : \mathfrak{g}_1 \times \mathfrak{g}_2 \rightarrow G_1 \times G_2, \quad (\xi_1, \xi_2) \mapsto (\exp_1(\xi_1), \exp_2(\xi_2)).$$

For example, in case of a n D vector space, or infinite-dimensional Banach space, the exponential map is the identity.

The unit circle in the complex plane $S^1 = \{z \in \mathbb{C} : |z| = 1\}$ is an Abelian Lie group under multiplication. The tangent space $T_e S^1$ is the imaginary

axis, and we identify \mathbb{R} with $T_e S^1$ by $t \mapsto 2\pi it$. With this identification, the exponential map $\exp : \mathbb{R} \rightarrow S^1$ is given by $\exp(t) = e^{2\pi it}$.

The n D torus $T^n = S^1 \times \cdots \times S^1$ (n times) is an Abelian Lie group. The exponential map $\exp : \mathbb{R}^n \rightarrow T^n$ is given by

$$\exp(t_1, \dots, t_n) = (e^{2\pi it_1}, \dots, e^{2\pi it_n}).$$

Since $S^1 = \mathbb{R}/\mathbb{Z}$, it follows that

$$T^n = \mathbb{R}^n / \mathbb{Z}^n,$$

the projection $\mathbb{R}^n \rightarrow T^n$ being given by the \exp map (see [Marsden and Ratiu (1999); Postnikov (1986)]).

For every $g \in G$, the map

$$Ad_g = T_e (R_{g^{-1}} \circ L_g) : \mathfrak{g} \rightarrow \mathfrak{g}$$

is called the *adjoint map* (or *operator*) associated with g .

For each $\xi \in \mathfrak{g}$ and $g \in G$ we have

$$\exp(Ad_g \xi) = g(\exp \xi)g^{-1}.$$

The relation between the adjoint map and the Lie bracket is the following: For all $\xi, \eta \in \mathfrak{g}$ we have

$$\left. \frac{d}{dt} \right|_{t=0} Ad_{\exp(t\xi)} \eta = [\xi, \eta].$$

A Lie subgroup H of G is a subgroup H of G which is also a submanifold of G . Then \mathfrak{h} is a Lie subalgebra of \mathfrak{g} and moreover $\mathfrak{h} = \{\xi \in \mathfrak{g} \mid \exp(t\xi) \in H, \text{ for all } t \in \mathbb{R}\}$.

Recall that one can characterize *Lebesgue measure* up to a multiplicative constant on \mathbb{R}^n by its invariance under translations. Similarly, on a locally compact group there is a unique (up to a nonzero multiplicative constant) left-invariant measure, called *Haar measure*. For Lie groups the existence of such measures is especially simple [Marsden and Ratiu (1999)]: Let G be a Lie group. Then there is a volume form $Ub5$, unique up to nonzero multiplicative constants, that is left-invariant. If G is compact, $Ub5$ is right invariant as well.

3.8.2 Actions of Lie Groups on Smooth Manifolds

Let M be a smooth manifold. An *action of a Lie group* G (with the unit element e) on M is a smooth map $\phi : G \times M \rightarrow M$, such that for all $x \in M$

and $g, h \in G$, (i) $\phi(e, x) = x$ and (ii) $\phi(g, \phi(h, x)) = \phi(gh, x)$. In other words, letting $\phi_g : x \in M \mapsto \phi_g(x) = \phi(g, x) \in M$, we have (i') $\phi_e = id_M$ and (ii') $\phi_g \circ \phi_h = \phi_{gh}$. ϕ_g is a diffeomorphism, since $(\phi_g)^{-1} = \phi_{g^{-1}}$. We say that the map $g \in G \mapsto \phi_g \in Diff(M)$ is a homomorphism of G into the group of diffeomorphisms of M . In case that M is a vector space and each ϕ_g is a linear operator, the function of G on M is called a representation of G on M [Puta (1993)]

An action ϕ of G on M is said to be *transitive group action*, if for every $x, y \in M$, there is $g \in G$ such that $\phi(g, x) = y$; *effective group action*, if $\phi_g = id_M$ implies $g = e$, that is $g \mapsto \phi_g$ is 1-1; and *free group action*, if for each $x \in M$, $g \mapsto \phi_g(x)$ is 1-1.

For example,

- (1) $G = \mathbb{R}$ acts on $M = \mathbb{R}$ by translations; explicitly,

$$\phi : G \times M \rightarrow M, \quad \phi(s, x) = x + s.$$

Then for $x \in \mathbb{R}$, $O_x = \mathbb{R}$. Hence M/G is a single point, and the action is transitive and free.

- (2) A complete flow ϕ_t of a vector-field X on M gives an action of \mathbb{R} on M , namely

$$(t, x) \in \mathbb{R} \times M \mapsto \phi_t(x) \in M.$$

- (3) Left translation $L_g : G \rightarrow G$ defines an effective action of G on itself. It is also transitive.
- (4) The coadjoint action of G on \mathfrak{g}^* is given by

$$Ad^* : (g, \alpha) \in G \times \mathfrak{g}^* \mapsto Ad_{g^{-1}}^*(\alpha) = (T_e(R_{g^{-1}} \circ L_g))^* \alpha \in \mathfrak{g}^*.$$

Let ϕ be an action of G on M . For $x \in M$ the *orbit* of x is defined by

$$O_x = \{\phi_g(x) | g \in G\} \subset M$$

and the *isotropy group* of ϕ at x is given by

$$G_x = \{g \in G | \phi(g, x) = x\} \subset G.$$

An action ϕ of G on a manifold M defines an equivalence relation on M by the relation belonging to the same orbit; explicitly, for $x, y \in M$, we write $x \sim y$ if there exists a $g \in G$ such that $\phi(g, x) = y$, that is, if $y \in O_x$. The set of all orbits M/G is called the *group orbit space*.

For example, let $M = \mathbb{R}^2 \setminus \{0\}$, $G = SO(2)$, the group of rotations in plane, and the action of G on M given by

$$\left(\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}, (x, y) \right) \mapsto (x \cos \theta - y \sin \theta, x \sin \theta + y \cos \theta).$$

The action is always free and effective, and the orbits are concentric circles, thus the orbit space is $M/G \simeq \mathbb{R}_+^*$.

A crucial concept in mechanics is the infinitesimal description of an action. Let $\phi : G \times M \rightarrow M$ be an action of a Lie group G on a smooth manifold M . For each $\xi \in \mathfrak{g}$,

$$\phi_\xi : \mathbb{R} \times M \rightarrow M, \quad \phi_\xi(t, x) = \phi(\exp(t\xi), x)$$

is an \mathbb{R} -action on M . Therefore, $\phi_{\exp(t\xi)} : M \rightarrow M$ is a flow on M ; the corresponding vector-field on M , given by

$$\xi_M(x) = \left. \frac{d}{dt} \right|_{t=0} \phi_{\exp(t\xi)}(x)$$

is called the infinitesimal generator of the action, corresponding to ξ in \mathfrak{g} .

The tangent space at x to an orbit O_x is given by

$$T_x O_x = \{\xi_M(x) | \xi \in \mathfrak{g}\}.$$

Let $\phi : G \times M \rightarrow M$ be a smooth G -action. For all $g \in G$, all $\xi, \eta \in \mathfrak{g}$ and all $\alpha, \beta \in \mathbb{R}$, we have:

$$(Ad_g \xi)_M = \phi_{g^{-1}}^* \xi_M, [\xi_M, \eta_M] = -[\xi, \eta]_M, \text{ and } (\alpha\xi + \beta\eta)_M = \alpha\xi_M + \beta\eta_M.$$

Let M be a smooth manifold, G a Lie group and $\phi : G \times M \rightarrow M$ a G -action on M . We say that a smooth map $f : M \rightarrow M$ is with respect to this action if for all $g \in G$,

$$f \circ \phi_g = \phi_g \circ f.$$

Let $f : M \rightarrow M$ be an equivariant smooth map. Then for any $\xi \in \mathfrak{g}$ we have

$$Tf \circ \xi_M = \xi_M \circ f.$$

3.8.3 Basic Dynamical Lie Groups

Here we give the first two examples of Lie groups, namely Galilei group and general linear group. Further examples will be given in association with particular dynamical systems.

3.8.3.1 Galilei Group

The *Galilei group* is the group of transformations in space and time that connect those Cartesian systems that are termed ‘inertial frames’ in Newtonian mechanics. The most general relationship between two such frames is the following. The origin of the time scale in the inertial frame S' may be shifted compared with that in S ; the orientation of the Cartesian axes in S' may be different from that in S ; the origin O of the Cartesian frame in S' may be moving relative to the origin O in S at a uniform velocity. The transition from S to S' involves ten parameters; thus the Galilei group is a ten parameter group. The basic assumption inherent in Galilei–Newtonian relativity is that there is an absolute time scale, so that the only way in which the time variables used by two different ‘inertial observers’ could possibly differ is that the zero of time for one of them may be shifted relative to the zero of time for the other.

Galilei space–time structure involves the following three elements:

- (1) *World*, as a 4D affine space A^4 . The points of A^4 are called *world points* or *events*. The parallel transitions of the world A^4 form a linear (i.e., Euclidean) space \mathbb{R}^4 .
- (2) *Time*, as a linear map $t : \mathbb{R}^4 \rightarrow \mathbb{R}$ of the linear space of the world parallel transitions onto the real ‘time axes’. Time interval from the event $a \in A^4$ to $b \in A^4$ is called the number $t(b-a)$; if $t(b-a) = 0$ then the events a and b are called synchronous. The set of all mutually synchronous events consists a 3D affine space A^3 , being a subspace of the world A^4 . The kernel of the mapping t consists of the parallel transitions of A^4 translating arbitrary (and every) event to the synchronous one; it is a linear 3D subspace \mathbb{R}^3 of the space \mathbb{R}^4 .
- (3) *Distance (metric)* between the synchronous events,

$$\rho(a, b) = \|a - b\|, \quad \text{for all } a, b \in A^3,$$

given by the scalar product in \mathbb{R}^3 . The distance transforms arbitrary space of synchronous events into the well known 3D Euclidean space E^3 .

The space A^4 , with the Galilei space–time structure on it, is called Galilei space. Galilei group is the group of all possible transformations of the Galilei space, preserving its structure. The elements of the Galilei group are called Galilei transformations. Therefore, Galilei transformations are affine transformations of the world A^4 preserving the time intervals and distances between the synchronous events.

The direct product $\mathbb{R} \times \mathbb{R}^3$, of the time axes with the 3D linear space \mathbb{R}^3 with a fixed Euclidean structure, has a natural Galilei structure. It is called Galilei coordinate system.

3.8.3.2 General Linear Group

The group of linear isomorphisms of \mathbb{R}^n to \mathbb{R}^n is a Lie group of dimension n^2 , called the *general linear group* and denoted $Gl(n, \mathbb{R})$. It is a smooth manifold, since it is a subset of the vector space $L(\mathbb{R}^n, \mathbb{R}^n)$ of all linear maps of \mathbb{R}^n to \mathbb{R}^n , as $Gl(n, \mathbb{R})$ is the inverse image of $\mathbb{R} \setminus \{0\}$ under the continuous map $A \mapsto \det A$ of $L(\mathbb{R}^n, \mathbb{R}^n)$ to \mathbb{R} . The group operation is composition

$$(A, B) \in Gl(n, \mathbb{R}) \times Gl(n, \mathbb{R}) \mapsto A \circ B \in Gl(n, \mathbb{R})$$

and the inverse map is

$$A \in Gl(n, \mathbb{R}) \mapsto A^{-1} \in Gl(n, \mathbb{R}).$$

If we choose a basis in \mathbb{R}^n , we can represent each element $A \in Gl(n, \mathbb{R})$ by an invertible $(n \times n)$ –matrix. The group operation is then matrix multiplication and the inversion is matrix inversion. The identity is the identity matrix I_n . The group operations are smooth since the formulas for the product and inverse of matrices are smooth in the matrix components.

The Lie algebra of $Gl(n, \mathbb{R})$ is $\mathfrak{gl}(n)$, the vector space $L(\mathbb{R}^n, \mathbb{R}^n)$ of all linear transformations of \mathbb{R}^n , with the commutator bracket

$$[A, B] = AB - BA.$$

For every $A \in L(\mathbb{R}^n, \mathbb{R}^n)$,

$$\gamma_A : t \in \mathbb{R} \mapsto \gamma_A(t) = \sum_{i=0}^{\infty} \frac{t^i}{i!} A^i \in Gl(n, \mathbb{R})$$

is a one–parameter subgroup of $Gl(n, \mathbb{R})$, because

$$\gamma_A(0) = I, \quad \text{and} \quad \dot{\gamma}_A(t) = \sum_{i=0}^{\infty} \frac{t^{i-1}}{(i-1)!} A^i = \gamma_A(t) A.$$

Hence γ_A is an integral curve of the left-invariant vector-field X_A . Therefore, the exponential map is given by

$$\exp : A \in L(\mathbb{R}^n, \mathbb{R}^n) \mapsto \exp(A) \equiv e^A = \gamma_A(1) = \sum_{i=0}^{\infty} \frac{A^i}{i!} \in Gl(n, \mathbb{R}).$$

For each $A \in Gl(n, \mathbb{R})$ the corresponding adjoint map

$$Ad_A : L(\mathbb{R}^n, \mathbb{R}^n) \rightarrow L(\mathbb{R}^n, \mathbb{R}^n)$$

is given by

$$Ad_AB = A \cdot B \cdot A^{-1}.$$

3.8.4 Application: Lie Groups in Biodynamics

3.8.4.1 Lie Groups of Joint Rotations

Recall (see [Ivancevic and Ivancevic (2006)]) that local kinematics at each rotational robot or (synovial) human joint, is defined as a *group action* of an n D constrained rotational Lie group $SO(n)$ on the Euclidean space \mathbb{R}^n . In particular, there is an action of $SO(2)$ —group in uniaxial human joints (cylindrical, or *hinge joints*, like knee and elbow) and an action of $SO(3)$ —group in three-axial human joints (spherical, or *ball-and-socket joints*, like hip, shoulder, neck, wrist and ankle). In both cases, $SO(n)$ acts, with its operators of rotation, on the vector $x = \{x^\mu\}$, ($i = 1, 2, 3$) of external, Cartesian coordinates of the parent body-segment, depending, at the same time, on the vector $q = \{q^s\}$, ($s = 1, \dots, n$) on n group-parameters, i.e., joint angles.

Each joint rotation $R \in SO(n)$ defines a map

$$R : x^\mu \mapsto \dot{x}^\mu, \quad R(x^\mu, q^s) = R_{q^s} x^\mu,$$

where $R_{q^s} \in SO(n)$ are joint group operators. The vector $v = \{v_s\}$, ($s = 1, \dots, n$) of n infinitesimal generators of these rotations, i.e., joint angular velocities, given by

$$v_s = -\left[\frac{\partial R(x^\mu, q^s)}{\partial q^s}\right]_{q=0} \frac{\partial}{\partial x^\mu},$$

constitute an n D Lie algebra $\mathfrak{so}(n)$ corresponding to the joint rotation group $SO(n)$. Conversely, each joint group operator R_{q^s} , representing a one-parameter subgroup of $SO(n)$, is defined as the exponential map of the

corresponding joint group generator v_s

$$R_{q^s} = \exp(q^s v_s). \quad (3.53)$$

The exponential map (3.53) represents a solution of the joint operator differential equation in the joint group-parameter space $\{q^s\}$

$$\frac{dR_{q^s}}{dq^s} = v_s R_{q^s}.$$

Uniaxial Group of Joint Rotations

The uniaxial joint rotation in a single Cartesian plane around a perpendicular axis, e.g., xy -plane about the z axis, by an internal joint angle θ , leads to the following transformation of the joint coordinates

$$\dot{x} = x \cos \theta - y \sin \theta, \quad \dot{y} = x \sin \theta + y \cos \theta.$$

In this way, the joint $SO(2)$ -group, given by

$$SO(2) = \left\{ R_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \mid \theta \in [0, 2\pi] \right\},$$

acts in a canonical way on the Euclidean plane \mathbb{R}^2 by

$$SO(2) = \left(\begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}, \begin{pmatrix} x \\ y \end{pmatrix} \right) \mapsto \begin{pmatrix} x \cos \theta - y \sin \theta \\ x \sin \theta + y \cos \theta \end{pmatrix}.$$

Its associated Lie algebra $\mathfrak{so}(2)$ is given by

$$\mathfrak{so}(2) = \left\{ \begin{pmatrix} 0 & -t \\ t & 0 \end{pmatrix} \mid t \in \mathbb{R} \right\},$$

since the curve $\gamma_\theta \in SO(2)$ given by

$$\gamma_\theta : t \in \mathbb{R} \mapsto \gamma_\theta(t) = \begin{pmatrix} \cos t\theta & -\sin t\theta \\ \sin t\theta & \cos t\theta \end{pmatrix} \in SO(2),$$

passes through the identity $I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and then

$$\left. \frac{d}{dt} \right|_{t=0} \gamma_\theta(t) = \begin{pmatrix} 0 & -\theta \\ \theta & 0 \end{pmatrix},$$

so that I_2 is a basis of $\mathfrak{so}(2)$, since $\dim(SO(2)) = 1$.

The *exponential map* $\exp : \mathfrak{so}(2) \rightarrow SO(2)$ is given by

$$\exp \begin{pmatrix} 0 & -\theta \\ \theta & 0 \end{pmatrix} = \gamma_\theta(1) = \begin{pmatrix} \cos t\theta & -\sin t\theta \\ \sin t\theta & \cos t\theta \end{pmatrix}.$$

The *infinitesimal generator* of the action of $SO(2)$ on \mathbb{R}^2 , i.e., joint angular velocity v , is given by

$$v = -y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y},$$

since

$$v_{\mathbb{R}^2}(x, y) = \left. \frac{d}{dt} \right|_{t=0} \exp(tv)(x, y) = \left. \frac{d}{dt} \right|_{t=0} \begin{pmatrix} \cos tv & -\sin tv \\ \sin tv & \cos tv \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

The *momentum map* (see subsection 3.12.3.5 below) $J : T^*\mathbb{R}^2 \rightarrow \mathbb{R}$ associated to the lifted action of $SO(2)$ on $T^*\mathbb{R}^2 \simeq \mathbb{R}^4$ is given by

$$\begin{aligned} J(x, y, p_1, p_2) &= xp_y - yp_x, \quad \text{since} \\ J(x, y, p_x, p_y)(\xi) &= (p_x dx + p_y dy)(v_{\mathbb{R}^2}) = -vp_x y + -vp_y x. \end{aligned}$$

The Lie group $SO(2)$ acts on the symplectic manifold $(\mathbb{R}^4, \omega = dp_x \wedge dx + dp_y \wedge dy)$ by

$$\begin{aligned} &\phi \left(\begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}, (x, y, p_x, p_y) \right) \\ &= (x \cos \theta - y \sin \theta, x \sin \theta + y \cos \theta, p_x \cos \theta - p_y \sin \theta, p_x \sin \theta + p_y \cos \theta). \end{aligned}$$

Three-Axial Group of Joint Rotations

The three-axial $SO(3)$ -group of human-like joint rotations depends on three parameters, Euler joint angles $q^i = (\varphi, \psi, \theta)$, defining the rotations about the Cartesian coordinate triad (x, y, z) placed at the joint pivot point. Each of the Euler angles are defined in the constrained range $(-\pi, \pi)$, so the joint group space is a constrained sphere of radius π .

Let $G = SO(3) = \{A \in \mathcal{M}_{3 \times 3}(\mathbb{R}) : A^t A = I_3, \det(A) = 1\}$ be the group of rotations in \mathbb{R}^3 . It is a Lie group and $\dim(G) = 3$. Let us isolate its one-parameter joint subgroups, i.e., consider the three operators of the finite joint rotations $R_\varphi, R_\psi, R_\theta \in SO(3)$, given by

$$R_\varphi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi & -\sin \varphi \\ 0 & \sin \varphi & \cos \varphi \end{bmatrix}, \quad R_\psi = \begin{bmatrix} \cos \psi & 0 & \sin \psi \\ 0 & 1 & 0 \\ -\sin \psi & 0 & \cos \psi \end{bmatrix}, \quad R_\theta = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

corresponding respectively to rotations about x -axis by an angle φ , about y -axis by an angle ψ , and about z -axis by an angle θ .

The total three-axial joint rotation A is defined as the product of above one-parameter rotations $R_\varphi, R_\psi, R_\theta$, i.e., $A = R_\varphi \cdot R_\psi \cdot R_\theta$ is equal

$$A = \begin{bmatrix} \cos \psi \cos \varphi - \cos \theta \sin \varphi \sin \psi & \cos \psi \cos \varphi + \cos \theta \cos \varphi \sin \psi & \sin \theta \sin \psi \\ -\sin \psi \cos \varphi - \cos \theta \sin \varphi \sin \psi & -\sin \psi \sin \varphi + \cos \theta \cos \varphi \cos \psi & \sin \theta \cos \psi \\ \sin \theta \sin \varphi & -\sin \theta \cos \varphi & \cos \theta \end{bmatrix}.$$

However, the order of these matrix products matters: different order products give different results, as the matrix product is *noncommutative product*. This is the reason why Hamilton's *quaternions*⁵ are today commonly used to parameterize the $SO(3)$ -group, especially in the field of 3D computer graphics.

The one-parameter rotations $R_\varphi, R_\psi, R_\theta$ define curves in $SO(3)$ starting from $I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Their derivatives in $\varphi = 0, \psi = 0$ and $\theta = 0$ belong to the associated *tangent Lie algebra* $\mathfrak{so}(3)$. That is the corresponding infinitesimal generators of joint rotations – joint angular velocities $v_\varphi, v_\psi, v_\theta \in \mathfrak{so}(3)$ – are respectively given by

$$\begin{aligned} v_\varphi &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} = -y \frac{\partial}{\partial z} + z \frac{\partial}{\partial y}, & v_\psi &= \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix} = -z \frac{\partial}{\partial x} + x \frac{\partial}{\partial z}, \\ v_\theta &= \begin{bmatrix} 0 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = -x \frac{\partial}{\partial y} + y \frac{\partial}{\partial x}. \end{aligned}$$

Moreover, the elements are linearly independent and so

$$\mathfrak{so}(3) = \left\{ \begin{bmatrix} 0 & -a & b \\ a & 0 & -\gamma \\ -b & \gamma & 0 \end{bmatrix} \mid a, b, \gamma \in \mathbb{R} \right\}.$$

The Lie algebra $\mathfrak{so}(3)$ is identified with \mathbb{R}^3 by associating to each $v =$

⁵Recall that the set of Hamilton's *quaternions* \mathbb{H} represents an extension of the set of complex numbers \mathbb{C} . We can compute a rotation about the unit vector, \mathbf{u} by an angle θ . The quaternion q that computes this rotation is

$$q = \left(\cos \frac{\theta}{2}, u \sin \frac{\theta}{2} \right).$$

$(v_\varphi, v_\psi, v_\theta) \in \mathbb{R}^3$ the matrix $v \in \mathfrak{so}(3)$ given by $v = \begin{bmatrix} 0 & -a & b \\ a & 0 & -\gamma \\ -b & \gamma & 0 \end{bmatrix}$. Then we have the following identities:

- (1) $\widehat{u \times v} = [\hat{u}, v]$; and
- (2) $u \cdot v = -\frac{1}{2} \text{Tr}(\hat{u} \cdot v)$.

The exponential map $\exp : \mathfrak{so}(3) \rightarrow SO(3)$ is given by *Rodrigues relation*

$$\exp(v) = I + \frac{\sin \|v\|}{\|v\|} v + \frac{1}{2} \left(\frac{\sin \frac{\|v\|}{2}}{\frac{\|v\|}{2}} \right)^2 v^2,$$

where the norm $\|v\|$ is given by

$$\|v\| = \sqrt{(v^1)^2 + (v^2)^2 + (v^3)^2}.$$

The dual, *cotangent Lie algebra* $\mathfrak{so}(3)^*$, includes the three joint angular momenta $p_\varphi, p_\psi, p_\theta \in \mathfrak{so}(3)^*$, derived from the joint velocities v by multiplying them with corresponding moments of inertia.

Note that the parameterization of $SO(3)$ -rotations is the subject of continuous research and development in many theoretical and applied fields of mechanics, such as rigid body, structural, and multi-body dynamics, robotics, spacecraft attitude dynamics, navigation, image processing, etc. For a complete discussion on the classical attitude representations see [Friberg (1988); Mladenova (1991); Shuster (1993); Schaub (1995)]. In addition, a modern vectorial parameterization of finite rotations, encompassing the mentioned earlier developments as well as Gibbs, Wiener, and Milenkovic parameterizations [Mladenova (1999); Bauchau and Trainelli (2003)].

3.8.4.2 Euclidean Groups of Total Joint Motions

Biodynamically realistic joint movement is predominantly rotational, plus restricted translational (translational motion in human joints is observed after reaching the limit of rotational amplitude). Gross translation in any human joint means joint dislocation, which is a severe injury. Obvious models for uniaxial and triaxial joint motions are *special Euclidean groups of rigid body motions*, $SE(2)$ and $SE(3)$, respectively.

Special Euclidean Group in the Plane

The motion in uniaxial human joints is naturally modelled by the *special Euclidean group in the plane*, $SE(2)$. It consists of all transformations of \mathbb{R}^2 of the form $Az + a$, where $z, a \in \mathbb{R}^2$, and

$$A \in SO(2) = \left\{ \text{matrices of the form } \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \right\}.$$

In other words [Marsden and Ratiu (1999)], group $SE(2)$ consists of matrices of the form:

$$(R_\theta, a) = \begin{pmatrix} R_\theta & a \\ 0 & I \end{pmatrix}, \text{ where } a \in \mathbb{R}^2 \text{ and } R_\theta \text{ is the rotation matrix:}$$

$$R_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}, \text{ while } I \text{ is the } 3 \times 3 \text{ identity matrix. The inverse } (R_\theta, a)^{-1} \text{ is given by}$$

$$(R_\theta, a)^{-1} = \begin{pmatrix} R_\theta & a \\ 0 & I \end{pmatrix}^{-1} = \begin{pmatrix} R_{-\theta} & -R_{-\theta}a \\ 0 & I \end{pmatrix}.$$

The Lie algebra $\mathfrak{se}(2)$ of $SE(2)$ consists of 3×3 block matrices of the form

$$\begin{pmatrix} -\xi J & v \\ 0 & 0 \end{pmatrix}, \quad \text{where} \quad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad (J^T = J^{-1} = -J),$$

with the usual commutator bracket. If we identify $\mathfrak{se}(2)$ with \mathbb{R}^3 by the isomorphism

$$\begin{pmatrix} -\xi J & v \\ 0 & 0 \end{pmatrix} \in \mathfrak{se}(2) \mapsto (\xi, v) \in \mathbb{R}^3,$$

then the expression for the Lie algebra bracket becomes

$$[(\xi, v_1, v_2), (\zeta, w_1, w_2)] = (0, \zeta v_2 - \xi w_2, \xi w_1 - \zeta v_1) = (0, \xi J^T w - \zeta J^T v),$$

where $v = (v_1, v_2)$ and $w = (w_1, w_2)$.

The *adjoint group action* of

$$(R_\theta, a) \begin{pmatrix} R_\theta & a \\ 0 & I \end{pmatrix} \quad \text{on} \quad (\xi, v) = \begin{pmatrix} -\xi J & v \\ 0 & 0 \end{pmatrix}$$

is given by the *group conjugation*,

$$\begin{pmatrix} R_\theta & a \\ 0 & I \end{pmatrix} \begin{pmatrix} -\xi J & v \\ 0 & 0 \end{pmatrix} \begin{pmatrix} R_{-\theta} & -R_{-\theta}a \\ 0 & I \end{pmatrix} = \begin{pmatrix} -\xi J & \xi Ja + R_\theta v \\ 0 & 0 \end{pmatrix},$$

or, in coordinates [Marsden and Ratiu (1999)],

$$Ad_{(R_\theta, a)}(\xi, v) = (\xi, \xi Ja + R_\theta v). \quad (3.54)$$

In proving (3.54) we used the identity $R_\theta J = JR_\theta$. Identify the dual algebra, $\mathfrak{se}(2)^*$, with matrices of the form $\begin{pmatrix} \frac{\mu}{2} J & 0 \\ \alpha & 0 \end{pmatrix}$, via the nondegenerate pairing given by the trace of the product. Thus, $\mathfrak{se}(2)^*$ is isomorphic to \mathbb{R}^3 via

$$\begin{pmatrix} \frac{\mu}{2} J & 0 \\ \alpha & 0 \end{pmatrix} \in \mathfrak{se}(2)^* \longmapsto (\mu, \alpha) \in \mathbb{R}^3,$$

so that in these coordinates, the pairing between $\mathfrak{se}(2)^*$ and $\mathfrak{se}(2)$ becomes

$$\langle (\mu, \alpha), (\xi, v) \rangle = \mu \xi + \alpha \cdot v,$$

that is, the usual dot product in \mathbb{R}^3 . The *coadjoint group action* is thus given by

$$Ad_{(R_\theta, a)}^*(\mu, \alpha) = (\mu - R_\theta \alpha \cdot Ja + R_\theta \alpha). \quad (3.55)$$

Formula (3.55) shows that the coadjoint orbits are the cylinders $T^*S_\alpha^1 = \{(\mu, \alpha) \mid \|\alpha\| = \text{const}\}$ if $\alpha \neq 0$ together with the points are on the μ -axis. The canonical cotangent bundle projection $\pi : T^*S_\alpha^1 \rightarrow S_\alpha^1$ is defined as $\pi(\mu, \alpha) = \alpha$.

Special Euclidean Group in the 3D Space

The most common group structure in human-like biodynamics is the *special Euclidean group in 3D space*, $SE(3)$. It is defined as a semidirect (noncommutative) product of 3D rotations and 3D translations, $SO(3) \triangleright \mathbb{R}^3$.

The Heavy Top

As a starting point consider a rigid body (see (3.12.3.2) below) moving with a fixed point but under the influence of gravity. This problem still has a configuration space $SO(3)$, but the symmetry group is only the circle group S^1 , consisting of rotations about the direction of gravity. One says that gravity has broken the symmetry from $SO(3)$ to S^1 . This time, eliminating the S^1 symmetry mysteriously leads one to the larger Euclidean group $SE(3)$ of rigid motion of \mathbb{R}^3 . Conversely, we can start with $SE(3)$ as

the configuration space for the rigid-body and ‘reduce out’ translations to arrive at $SO(3)$ as the configuration space (see [Marsden and Ratiu (1999)]).

The equations of motion for a rigid body with a fixed point in a gravitational field give an interesting example of a system that is Hamiltonian (see (3.12.3.2)) relative to a *Lie–Poisson bracket* (see (3.13.2)). The underlying Lie algebra consists of the algebra of infinitesimal Euclidean motions in \mathbb{R}^3 .

The basic phase-space we start with is again $T^*SO(3)$, parameterized by Euler angles and their conjugate momenta. In these variables, the equations are in canonical Hamiltonian form. However, the presence of gravity breaks the symmetry, and the system is no longer $SO(3)$ invariant, so it cannot be written entirely in terms of the body angular momentum p . One also needs to keep track of Γ , the ‘direction of gravity’ as seen from the body. This is defined by $\Gamma = A^{-1}k$, where k points upward and A is the element of $SO(3)$ describing the current configuration of the body. The equations of motion are

$$\begin{aligned}\dot{p}_1 &= \frac{I_2 - I_3}{I_2 I_3} p_2 p_3 + Mgl(\Gamma^2 \chi^3 - \Gamma^3 \chi^2), \\ \dot{p}_2 &= \frac{I_3 - I_1}{I_3 I_1} p_3 p_1 + Mgl(\Gamma^3 \chi^1 - \Gamma^1 \chi^3), \\ \dot{p}_3 &= \frac{I_1 - I_2}{I_1 I_2} p_1 p_2 + Mgl(\Gamma^1 \chi^2 - \Gamma^2 \chi^1), \\ \text{and} \quad \dot{\Gamma} &= \Gamma \times \Omega,\end{aligned}$$

where Ω is the body’s angular velocity vector, I_1, I_2, I_3 are the body’s principal moments of inertia, M is the body’s mass, g is the acceleration of gravity, χ is the body fixed unit vector on the line segment connecting the fixed point with the body’s center of mass, and l is the length of this segment.

The Euclidean Group and Its Lie Algebra

An element of $SE(3)$ is a pair (A, a) where $A \in SO(3)$ and $a \in \mathbb{R}^3$. The action of $SE(3)$ on \mathbb{R}^3 is the rotation A followed by translation by the vector a and has the expression

$$(A, a) \cdot x = Ax + a.$$

Using this formula, one sees that multiplication and inversion in $SE(3)$ are given by

$$(A, a)(B, b) = (AB, Ab + a) \quad \text{and} \quad (A, a)^{-1} = (A^{-1}, -A^{-1}a),$$

for $A, B \in SO(3)$ and $a, b \in \mathbb{R}^3$. The identity element is $(I, 0)$.

The Lie algebra of the Euclidean group $SE(3)$ is $\mathfrak{se}(3) = \mathbb{R}^3 \times \mathbb{R}^3$ with the Lie bracket

$$[(\xi, u), (\eta, v)] = (\xi \times \eta, \xi \times v - \eta \times u). \quad (3.56)$$

The Lie algebra of the Euclidean group has a structure that is a special case of what is called a *semidirect product*. Here it is the *product of the group of rotations with the corresponding group of translations*. It turns out that semidirect products occur under rather general circumstances when the symmetry in T^*G is broken.

The dual Lie algebra of the Euclidean group $SE(3)$ is $\mathfrak{se}(3)^* = \mathbb{R}^3 \times \mathbb{R}^3$ with the same Lie bracket (3.56). For the further details on adjoint orbits in $\mathfrak{se}(3)$ as well as coadjoint orbits in $\mathfrak{se}(3)^*$ see [Marsden and Ratiu (1999)].

Symplectic Group in Hamiltonian Mechanics

Let $J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$, with I the $n \times n$ identity matrix. Now, $A \in L(\mathbb{R}^{2n}, \mathbb{R}^{2n})$ is called a *symplectic matrix* if $A^T J A = J$. Let $Sp(2n, \mathbb{R})$ be the set of $2n \times 2n$ symplectic matrices. Taking determinants of the condition $A^T J A = J$ gives $\det A = \pm 1$, and so $A \in GL(2n, \mathbb{R})$. Furthermore, if $A, B \in Sp(2n, \mathbb{R})$, then $(AB)^T J (AB) = B^T A^T J A B = J$. Hence, $AB \in Sp(2n, \mathbb{R})$, and if $A^T J A = J$, then $JA = (A^T)^{-1} J = (A^{-1})^T J$, so $J = (A - 1)^T J A^{-1}$, or $A^{-1} \in Sp(2n, \mathbb{R})$. Thus, $Sp(2n, \mathbb{R})$ is a group [Marsden and Ratiu (1999)].

The *symplectic Lie group*

$$Sp(2n, \mathbb{R}) = \{A \in GL(2n, \mathbb{R}) : A^T J A = J\}$$

is a noncompact, connected Lie group of dimension $2n^2 + n$. Its Lie algebra

$$\mathfrak{sp}(2n, \mathbb{R}) = \{A \in L(\mathbb{R}^{2n}, \mathbb{R}^{2n}) : A^T J A = J = 0\},$$

called the *symplectic Lie algebra*, consists of the $2n \times 2n$ matrices A satisfying $A^T J A = 0$ [Marsden and Ratiu (1999)].

Consider a particle of mass m moving in a potential $V(q)$, where $q^i = (q^1, q^2, q^3) \in \mathbb{R}^3$. Newtonian second law states that the particle moves along a curve $q(t)$ in \mathbb{R}^3 in such a way that $m\ddot{q}^i = -\text{grad } V(q^i)$. Introduce the

momentum $p_i = m\dot{q}^i$, and the energy

$$H(q, p) = \frac{1}{2m} \sum_{i=1}^3 p_i^2 + V(q).$$

Then

$$\begin{aligned} \frac{\partial H}{\partial q^i} &= \frac{\partial V}{\partial q^i} = -m\ddot{q}^i = -\dot{p}_i, \quad \text{and} \\ \frac{\partial H}{\partial p_i} &= \frac{1}{m} p_i = \dot{q}^i, \quad (i = 1, 2, 3), \end{aligned}$$

and hence Newtonian law $F = m\ddot{q}^i$ is equivalent to Hamiltonian equations

$$\dot{q}^i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q^i}.$$

Now, writing $z = (q^i, p_i)$ [Marsden and Ratiu (1999)],

$$J \operatorname{grad} H(z) = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \begin{pmatrix} \frac{\partial H}{\partial q^i} \\ \frac{\partial H}{\partial p_i} \end{pmatrix} = (\dot{q}^i, \dot{p}_i) = \dot{z},$$

so Hamiltonian equations read

$$\dot{z} = J \operatorname{grad} H(z). \quad (3.57)$$

Now let $f : \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}^3 \times \mathbb{R}^3$ and write $w = f(z)$. If $z(t)$ satisfies Hamiltonian equations (3.57) then $w(t) = f(z(t))$ satisfies $\dot{w} = A^T \dot{z}$, where $A^T = [\partial w^i / \partial z^j]$ is the Jacobian matrix of f . By the chain rule,

$$\dot{w} = A^T \underset{z}{J \operatorname{grad} H(z)} = A^T \underset{w}{J \operatorname{grad} H(z(w))}.$$

Thus, the equations for $w(t)$ have the form of Hamiltonian equations with energy $K(w) = H(z(w))$ iff $A^T J A = J$, that is, iff A is symplectic. A nonlinear transformation f is canonical iff its Jacobian matrix is symplectic. $Sp(2n, \mathbb{R})$ is the linear invariance group of classical mechanics [Marsden and Ratiu (1999)].

3.8.4.3 Group Structure of Biodynamical Manifold

Purely Rotational Biodynamical Manifold

Kinematics of an n –segment human–body chain (like arm, leg or spine) is usually defined as a map between *external coordinates* (usually, end–effector coordinates) x^r ($r = 1, \dots, n$) and *internal joint coordinates* q^i ($i = 1, \dots, N$) (see [Ivancevic and Snoswell (2001); Ivancevic (2002); Ivancevic and Pearce (2001b); Ivancevic and Pearce (2001b); Ivancevic (2005)]). The *forward kinematics* are defined as a nonlinear map $x^r = x^r(q^i)$ with a corresponding linear vector functions $dx^r = \partial x^r / \partial q^i dq^i$ of differentials: and $\dot{x}^r = \partial x^r / \partial q^i \dot{q}^i$ of velocities. When the rank of the configuration–dependent Jacobian matrix $J \equiv \partial x^r / \partial q^i$ is less than n the *kinematic singularities* occur; the onset of this condition could be detected by the *manipulability measure*. The *inverse kinematics* are defined conversely by a nonlinear map $q^i = q^i(x^r)$ with a corresponding linear vector functions $dq^i = \partial q^i / \partial x^r dx^r$ of differentials and $\dot{q}^i = \partial q^i / \partial x^r \dot{x}^r$ of velocities. Again, in the case of *redundancy* ($n < N$), the inverse kinematic problem admits infinite solutions; often the *pseudo–inverse* configuration–control is used instead: $\dot{q}^i = J^* \dot{x}^r$, where $J^* = J^T(JJ^T)^{-1}$ denotes the Moore–Penrose pseudo–inverse of the Jacobian matrix J .

Humanoid joints, that is, internal coordinates q^i ($i = 1, \dots, N$), constitute a smooth configuration manifold M , described as follows. Uniaxial, ‘hinge’ joints represent constrained, rotational Lie groups $SO(2)_{cnstr}^i$, parameterized by constrained angles $q_{cnstr}^i \equiv q^i \in [q_{\min}^i, q_{\max}^i]$. Three–axial, ‘ball–and–socket’ joints represent constrained rotational Lie groups $SO(3)_{cnstr}^i$, parameterized by constrained Euler angles $q^i = q_{cnstr}^{\phi_i}$ (in the following text, the subscript ‘cnstr’ will be omitted, for the sake of simplicity, and always assumed in relation to internal coordinates q^i).

All $SO(n)$ –joints are Hausdorff C^∞ –manifolds with atlases (U_α, u_α) ; in other words, they are paracompact and metrizable smooth manifolds, admitting Riemannian metric.

Let A and B be two smooth manifolds described by smooth atlases (U_α, u_α) and (V_β, v_β) , respectively. Then the family $(U_\alpha \times V_\beta, u_\alpha \times v_\beta : U_\alpha \times V_\beta \rightarrow \mathbb{R}^m \times \mathbb{R}^n)_{(\alpha, \beta) \in A \times B}$ is a smooth atlas for the direct product $A \times B$. Now, if A and B are two Lie groups (say, $SO(n)$), then their *direct product* $G = A \times B$ is at the same time their direct product as smooth manifolds and their direct product as algebraic groups, with the

Reduction of the Rotational Biodynamical Manifold

The biodynamical configuration manifold M (Figure 3.6) can be (for the sake of the brain-like motor control) reduced to N —torus T^N , in three steps, as follows.

First, a single three-axial $SO(3)$ —joint can be reduced to the direct product of three uniaxial $SO(2)$ —joints, in the sense that three hinge joints can produce any orientation in space, just as a ball-joint can. Algebraically, this means reduction (using symbol ‘ \gtrsim ’) of each of the three $SO(3)$ rotation matrices to the corresponding $SO(2)$ rotation matrices

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & -\sin \phi \\ 0 & \sin \phi & \cos \phi \end{pmatrix} \gtrsim \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix}$$

$$\begin{pmatrix} \cos \psi & 0 & \sin \psi \\ 0 & 1 & 0 \\ -\sin \psi & 0 & \cos \psi \end{pmatrix} \gtrsim \begin{pmatrix} \cos \psi & \sin \psi \\ -\sin \psi & \cos \psi \end{pmatrix}$$

$$\begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \gtrsim \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

In this way we can set the *reduction equivalence relation* $SO(3) \gtrsim SO(2) \triangleright SO(2) \triangleright SO(2)$, where ‘ \triangleright ’ denotes the noncommutative *semidirect product* (see (3.8.4.2) above).

Second, we have a homeomorphism: $SO(2) \sim S^1$, where S^1 denotes the constrained unit circle in the complex plane, which is an Abelian Lie group.

Third, let I^N be the unit cube $[0, 1]^N$ in \mathbb{R}^N and ‘ \sim ’ an equivalence relation on \mathbb{R}^N get by ‘gluing’ together the opposite sides of I^N , preserving their orientation. The manifold of human-body configurations (Figure 3.6) can be represented as the *quotient space* of \mathbb{R}^N by the space of the integral lattice points in \mathbb{R}^N , that is a constrained ND torus T^N (3.221),

$$\mathbb{R}^N / Z^N = I^N / \sim \cong \prod_{i=1}^N S_i^1 \equiv \{(q^i, i = 1, \dots, N) : \text{mod } 2\pi\} = T^N. \quad (3.58)$$

Since S^1 is an Abelian Lie group, its N —fold tensor product T^N is also an Abelian Lie group, the toral group, of all nondegenerate diagonal $N \times N$ matrices. As a Lie group, the biodynamical configuration space $M \equiv T^N$

has a natural Banach manifold structure with local internal coordinates $q^i \in U$, U being an open set (chart) in T^N .

Conversely by ‘ungluing’ the configuration space we get the primary unit cube. Let ‘ \sim^* ’ denote an equivalent decomposition or ‘ungluing’ relation. By the Tychonoff product–topology Theorem, for every such quotient space there exists a ‘selector’ such that their quotient models are homeomorphic, that is, $T^N / \sim^* \approx A^N / \sim^*$. Therefore I^N represents a ‘selector’ for the configuration torus T^N and can be used as an N –directional ‘command–space’ for the topological control of human motion. Any subset of DOF on the configuration torus T^N representing the joints included in human motion has its simple, rectangular image in the command space – selector I^N . Operationally, this resembles what the *brain–motor–controller*, the *cerebellum*, actually performs on the highest level of human motor control.

The Complete Biodynamical Manifold

The full kinematics of a whole human–like body can be split down into *five kinematic chains*: one for each leg and arm, plus one for spine with the head. In all five chains internal joint coordinates, namely n_1 constrained rotations x_{rt}^k together with n_2 of even more constrained translations x_{tr}^j (see Figure 3.7), constitute a smooth nD anthropomorphic configuration manifold M , with local coordinates x^i , ($i = 1, \dots, n$). That is, the motion space in each joint is defined as a semidirect (noncommutative) product of the Lie group $SO(n)$ of constrained rotations and a corresponding Lie group \mathbb{R}^n of even more restricted translations. More precisely, in each movable human–like joint we have an action of the constrained special Euclidean $SE(3)$ group (see (3.8.4.2) above). The joints themselves are linked by direct (commutative) products.

Realistic Human Spine Manifold

The high–resolution human spine manifold is a dynamical chain consisting of 25 constrained $SE(3)$ – joints. Each movable spinal joint has 6 DOF: 3 dominant rotations, (performed first in any free spinal movement), restricted to about 7 angular degrees and 3 secondary translations (performed after reaching the limit of rotational amplitude), restricted to about 5 mm (see Figure 3.8).

Now, $SE(3) = SO(3) \triangleright \mathbb{R}^3$ is a non–compact group, so there is no natural metric given by the kinetic energy on $SE(3)$, and consequently, no

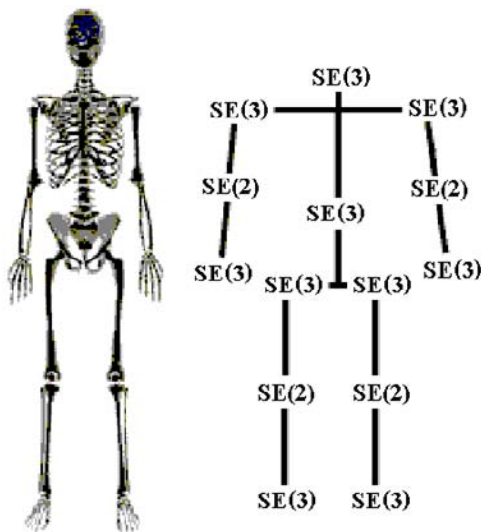


Fig. 3.7 A medium-resolution, whole-body biodynamical manifold, with just a single $SE(3)$ -joint representing the spinal movability.

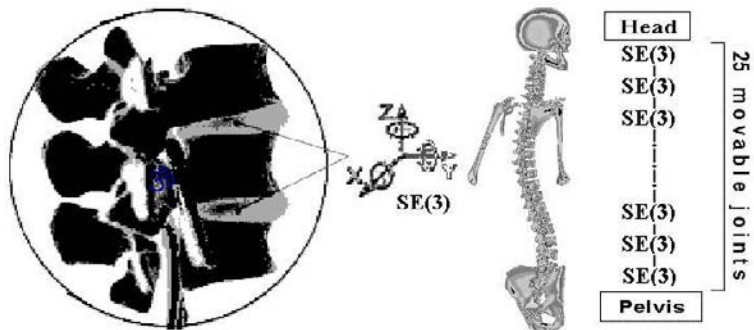


Fig. 3.8 The high-resolution human spine manifold is a dynamical chain consisting of 25 constrained $SE(3)$ -joints.

natural controls in the sense of geodesics on $SE(3)$. However, both of its subgroups, $SO(3)$ and \mathbb{R}^3 , are compact with quadratic metric forms defined by standard line element $g_{ij}dq^i dq^j$, and therefore admit optimal muscular-like controls in the sense of geodesics (see section 3.10.1.2 below).

3.8.5 Application: *Dynamical Games on $SE(n)$ –Groups*

In this section we propose a general approach to modelling *conflict resolution manoeuvres* for land, sea and airborne *vehicles*, using dynamical games on Lie groups. We use the generic name ‘vehicle’ to represent all planar vehicles, namely land and sea vehicles, as well as fixed altitude motion of aircrafts (see, e.g., [Lygeros *et. al.* (1998); Tomlin *et. al.* (1998)]). First, we elaborate on the two-vehicle conflict resolution manoeuvres, and after that discuss the multi-vehicle manoeuvres.

We explore special features of the dynamical games solution when the underlying dynamics correspond to left-invariant control systems on Lie groups. We show that the 2D (i.e., planar) motion of a vehicle may be modelled as a control system on the Lie group $SE(2)$. The proposed algorithm surrounds each vehicle with a circular protected zone, while the simplification in the derivation of *saddle* and *Nash strategies* follows from the use of symplectic reduction techniques [Marsden and Ratiu (1999)]. To model the two-vehicle conflict resolution, we construct the safe subset of the state-space for one of the vehicles using zero-sum non-cooperative dynamic game theory [Basar and Olsder (1995)] which we specialize to the $SE(2)$ group. If the underlying continuous dynamics are left-invariant control systems, reduction techniques can be used in the computation of safe sets.

3.8.5.1 Configuration Models for Planar Vehicles

The configuration of each individual vehicle is described by an element of the *Lie group* $SE(2)$ of rigid-body motions in \mathbb{R}^2 . Let $g_i \in SE(2)$ denote the configurations of vehicles labelled i , with $i = 1, 2$. The motion of each vehicle may be modelled as a left-invariant vector-field on $SE(2)$:

$$\dot{g}_i = g_i X_i, \quad (3.59)$$

where the vector-fields X_i belong to the vector space $\mathfrak{se}(2)$, the *Lie algebra* associated with the group $SE(2)$.

Each $g_i \in SE(2)$ can be represented in standard local coordinates (x_i, y_i, θ_i) as

$$g_i = \begin{bmatrix} \cos \theta_i & -\sin \theta_i & x_i \\ \sin \theta_i & \cos \theta_i & y_i \\ 0 & 0 & 1 \end{bmatrix},$$

where x_i, y_i is the position of vehicle i and θ_i is its orientation, or heading. The associated Lie algebra is $\mathfrak{se}(2)$, with $X_i \in \mathfrak{se}(2)$ represented as

$$X_i = \begin{bmatrix} 0 & -\omega_i & v_i \\ \omega_i & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

where v_i and ω_i represent the *translational* (linear) and *rotational* (angular) *velocities*, respectively.

Now, to determine *dynamics of the relative configuration* of two vehicles, we perform a change (transformation) of coordinates, to place the identity element of the group $SE(2)$ on vehicle 1. If $g^{rel} \in SE(2)$ denotes the relative configuration of vehicle 2 with respect to vehicle 1, then

$$g_2 = g_1 g^{rel} \implies g^{rel} = g_1^{-1} g_2.$$

Differentiation with respect to time yields the dynamics of the relative configuration:

$$\dot{g}^{rel} = g^{rel} X_2 - X_1 g^{rel},$$

which expands into:

$$\begin{aligned} \dot{x}^{rel} &= -v_1 + v_2 \cos \theta^{rel} + \omega_1 y^{rel}, \\ \dot{y}^{rel} &= v_2 \sin \theta^{rel} - \omega_1 x^{rel}, \\ \dot{\theta}^{rel} &= \omega_2 - \omega_1. \end{aligned}$$

3.8.5.2 Two-Vehicles Conflict Resolution Manoeuvres

Next, we seek *control strategies* for each vehicle, which are *safe* under (possible) *uncertainty* in the actions of neighbouring vehicle. For this, we expand the dynamics of two vehicles (3.59),

$$\dot{g}_1 = g_1 X_1, \quad \dot{g}_2 = g_2 X_2,$$

and write it in the matrix form as

$$\dot{g} = gX, \tag{3.60}$$

with

$$g = \begin{bmatrix} g_1 & 0 \\ 0 & g_2 \end{bmatrix}, \quad X = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix},$$

in which g is an element in the configuration manifold $M = SE(2) \times SE(2)$, while the vector-fields $X_i \in \mathfrak{se}(2) \times \mathfrak{se}(2)$ are linearly parameterised by velocity inputs $(\omega_1, v_1) \in \mathbb{R}^2$ and $(\omega_2, v_2) \in \mathbb{R}^2$.

The goal of each vehicle is to maintain *safe operation*, meaning that

(i) the vehicles remain outside of a specified *target set* T with boundary ∂T , defined by

$$T = \{g \in M | l(g) < 0\},$$

where $l(g)$ is a differentiable circular function,

$$l(g) = (x_2 - x_1)^2 + (y_2 - y_1)^2 - \rho^2$$

(with ρ denoting the *radius of a circular protected zone*) defines the minimum allowable lateral separation between vehicles; and

(ii)

$$dl(g) \neq 0 \quad \text{on} \quad \partial T = \{g \in M | l(g) = 0\},$$

where d represents the *exterior derivative* (a unique generalization of the gradient, divergence and curl).

Now, due to possible uncertainty in the actions of vehicle 2, the safest possible strategy of vehicle 1 is to drive along a *trajectory* which guarantees that the minimum allowable separation with vehicle 2 is maintained regardless of the actions of vehicle 2. We therefore formulate this problem as a *zero-sum dynamical game* with two players: control vs. disturbance. The *control* is the action of vehicle 1,

$$u = (\omega_1, v_1) \in U,$$

and the *disturbance* is the action of vehicle 2,

$$d = (\omega_2, v_2) \in D.$$

Here the control and disturbance sets, U and D , are defined as

$$\begin{aligned} U &= ([\omega_1^{\min}, \omega_1^{\max}], [v_1^{\min}, v_1^{\max}]), \\ D &= ([\omega_2^{\min}, \omega_2^{\max}], [v_2^{\min}, v_2^{\max}]) \end{aligned}$$

and the corresponding control and disturbance functional spaces, \mathcal{U} and \mathcal{D} are defined as:

$$\begin{aligned} \mathcal{U} &= \{u(\cdot) \in PC^0(\mathbb{R}^2) | u(t) \in U, t \in \mathbb{R}\}, \\ \mathcal{D} &= \{d(\cdot) \in PC^0(\mathbb{R}^2) | d(t) \in D, t \in \mathbb{R}\}, \end{aligned}$$

where $PC^0(\mathbb{R}^2)$ is the space of piecewise continuous functions over \mathbb{R}^2 .

We define the *cost of a trajectory* $g(t)$ which starts at state g at initial time $t \leq 0$, evolves according to (3.60) with input $(u(\cdot), d(\cdot))$, and ends at the final state $g(0)$ as:

$$J(g, u(\cdot), d(\cdot), t) : SE(2) \times SE(2) \times \mathcal{U} \times \mathcal{D} \times \mathbb{R}_- \rightarrow \mathbb{R},$$

such that $J(g, u(\cdot), d(\cdot), t) = l(g(0))$,

(3.61)

where 0 is the final time (without loss of generality). Thus the cost depends only on the final state $g(0)$ (the Lagrangian, or running cost, is identically zero). The game is won by vehicle 1 if the terminal state $g(0)$ is either outside T or on ∂T (i.e., $J(g, 0) \geq 0$), and is won by vehicle 2 otherwise.

This two-player zero-sum dynamical game on $SE(2)$ is defined as follows. Consider the matrix system (3.60), $\dot{g} = gX$, over the time interval $[t, 0]$ where $t < 0$ with the cost function $J(g, u(\cdot), d(\cdot), t)$ defined by (3.61) As vehicle 1 attempts to maximize this cost assuming that vehicle 2 is acting blindly, the optimal control action and worst disturbance actions are calculated as

$$u^* = \arg \max_{u \in \mathcal{U}} \min_{d \in \mathcal{D}} J(g, u(\cdot), d(\cdot), t), \quad d^* = \arg \min_{d \in \mathcal{D}} \max_{u \in \mathcal{U}} J(g, u(\cdot), d(\cdot), t).$$

The game is said to have a *saddle solution* (u^*, d^*) if the resulting optimal cost $J^*(g, t)$ does not depend on the order of play, i.e., on the order in which the maximization and minimization is performed:

$$J^*(g, t) = \max_{u \in \mathcal{U}} \min_{d \in \mathcal{D}} J(g, u(\cdot), d(\cdot), t) = \min_{d \in \mathcal{D}} \max_{u \in \mathcal{U}} J(g, u(\cdot), d(\cdot), t).$$

Using this saddle solution we calculate the ‘losing states’ for vehicle 1, called the *predecessor* $Pre_t(T)$ of the target set T ,

$$Pre_t(T) = \{g \in M \mid J(g, u^*(\cdot), d(\cdot), t) < 0\}.$$

3.8.5.3 Symplectic Reduction and Dynamical Games on $SE(2)$

Since vehicles 1 and 2 have dynamics given by left-invariant control systems on the Lie group $SE(2)$, we have

$$X_1 = \xi^1 \omega_1 + \xi^2 v_1, \quad X_2 = \xi^1 \omega_2 + \xi^2 v_2,$$

with ξ^1, ξ^2 being two of the three basis elements for the tangent Lie algebra $\mathfrak{se}(2)$ given by

$$\xi^1 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \xi^2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \xi^3 = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

If p_1 (resp. p_2) is a cotangent vector-field to $SE(2)$ at g_1 (resp. g_2), belonging to the cotangent (dual) Lie algebra $\mathfrak{se}(2)^*$, we can define the *momentum functions* for both vehicles:

$$\begin{aligned} P_1^1 &= \langle p_1, g_1 \xi^1 \rangle, P_1^2 = \langle p_1, g_1 \xi^2 \rangle, P_1^3 = \langle p_1, g_1 \xi^3 \rangle, \\ P_2^1 &= \langle p_2, g_2 \xi^1 \rangle, P_2^2 = \langle p_2, g_2 \xi^2 \rangle, P_2^3 = \langle p_2, g_2 \xi^3 \rangle, \end{aligned}$$

which can be compactly written as

$$P_i^j = \langle p_i, g_i \xi^j \rangle.$$

Defining $p = (p_1, p_2) \in \mathfrak{se}(2)^* \times \mathfrak{se}(2)^*$, the *optimal cost* for the two-player, zero-sum dynamical game is given by

$$J^*(g, t) = \max_{u \in \mathcal{U}} \min_{d \in \mathcal{D}} J(g, u(\cdot), d(\cdot), t) = \max_{u \in \mathcal{U}} \min_{d \in \mathcal{D}} l(g(0)).$$

The *Hamiltonian* $H(g, p, u, d)$ is given by

$$H(g, p, u, d) = P_1^1 \omega_1 + P_1^2 v_1 + P_2^1 \omega_1 + P_2^2 v_1$$

for control and disturbance inputs $(\omega_1, v_1) \in U$ and $(\omega_2, v_2) \in D$ as defined above. It follows that the *optimal Hamiltonian* $H^*(g, p)$, defined on the cotangent bundle $T^*SE(2)$, is given by

$$\begin{aligned} H^*(g, p) &= P_1^1 \frac{\omega_1^{\max} + \omega_1^{\min}}{2} + P_2^1 \frac{\omega_2^{\max} + \omega_2^{\min}}{2} + |P_1^1| \frac{\omega_1^{\max} - \omega_1^{\min}}{2} \\ &\quad - |P_1^1| \frac{\omega_2^{\max} - \omega_2^{\min}}{2} + P_1^2 \frac{v_1^{\max} + v_1^{\min}}{2} + P_2^2 \frac{v_2^{\max} + v_2^{\min}}{2} \\ &\quad + |P_1^2| \frac{v_1^{\max} - v_1^{\min}}{2} - |P_1^2| \frac{v_2^{\max} - v_2^{\min}}{2} \end{aligned}$$

and the *saddle solution* (u^*, d^*) is given by

$$u^* = \arg \max_{u \in \mathcal{U}} \min_{d \in \mathcal{D}} H(g, p, u, d), \quad d^* = \arg \min_{d \in \mathcal{D}} \max_{u \in \mathcal{U}} H(g, p, u, d). \quad (3.62)$$

Note that $H(g, p, u, d)$ and $H^*(g, p)$ do not depend on the state g and costate p directly, rather through the momentum functions P_1^j, P_2^j . This is because the dynamics are determined by left-invariant vector-fields on

the Lie group and the Lagrangian is state independent [Marsden and Ratiu (1999)].

The optimal Hamiltonian $H^*(g, p)$ determines a 12D Hamiltonian vector-field X_{H^*} on the symplectic manifold $T^*M = SE(2) \times SE(2) \times \mathfrak{se}(2)^* \times \mathfrak{se}(2)^*$ (which is the cotangent bundle of the configuration manifold M), defined by Hamiltonian equations

$$X_{H^*} : \dot{g} = \frac{\partial H^*(g, p)}{\partial p}, \quad \dot{p} = -\frac{\partial H^*(g, p)}{\partial g},$$

with initial condition at time t being $g(t) = g$ and final condition at time 0 being $p(0) = dl(g(0))$. In general, to solve for the saddle solution (3.62), one needs to solve the ODE system for all states. However since the original system on $M = SE(2) \times SE(2)$ is left-invariant, it induces generic symmetries in the Hamiltonian dynamics on T^*M , referred to as Marsden–Weinstein reduction of Hamiltonian systems on symplectic manifolds, see [Marsden and Ratiu (1999)]. In general for such systems one only needs to solve an ODE system with half of the dimensions of the underlying symplectic manifold.

For the two-vehicle case we only need to solve an ODE system with 6 states. That is exactly given by the dynamics of the 6 momentum functions

$$\dot{P}_i^j = L_{X_{H^*}} P_i^j = \{P_i^j, H^*(g, p)\}, \quad (3.63)$$

for $i, j = 1, 2$, which is the *Lie derivative* of P_i^j with respect to the Hamiltonian vector-field X_{H^*} . In the equation (3.63), the bracket $\{\cdot, \cdot\}$ is the *Poisson bracket* [Ivancevic and Pearce (2001a)], giving the commutation relations:

$$\begin{aligned} \{P_1^1, P_1^2\} &= P_1^3, & \{P_1^2, P_1^3\} &= 0, & \{P_1^3, P_1^1\} &= P_1^2, \\ \{P_2^1, P_2^2\} &= P_2^3, & \{P_2^2, P_2^3\} &= 0, & \{P_2^3, P_2^1\} &= P_2^2. \end{aligned}$$

Using these commutation relations, equation (3.63) can be written explic-

itly:

$$\begin{aligned}
 \dot{P}_1^1 &= P_1^3 \left(\frac{v_1^{\max} + v_1^{\min}}{2} + \text{sign}(P_1^2) \frac{v_1^{\max} - v_1^{\min}}{2} \right), \\
 \dot{P}_1^2 &= P_1^3 \left(-\frac{\omega_1^{\max} + \omega_1^{\min}}{2} - \text{sign}(P_1^1) \frac{\omega_1^{\max} - \omega_1^{\min}}{2} \right), \\
 \dot{P}_1^3 &= P_1^2 \left(\frac{\omega_1^{\max} + \omega_1^{\min}}{2} + \text{sign}(P_1^1) \frac{\omega_1^{\max} - \omega_1^{\min}}{2} \right), \\
 \dot{P}_2^1 &= P_2^3 \left(\frac{v_2^{\max} + v_2^{\min}}{2} + \text{sign}(P_2^2) \frac{v_2^{\max} - v_2^{\min}}{2} \right), \\
 \dot{P}_2^2 &= P_2^3 \left(-\frac{\omega_2^{\max} + \omega_2^{\min}}{2} - \text{sign}(P_2^1) \frac{\omega_2^{\max} - \omega_2^{\min}}{2} \right), \\
 \dot{P}_2^3 &= P_2^2 \left(\frac{\omega_2^{\max} + \omega_2^{\min}}{2} + \text{sign}(P_2^1) \frac{\omega_2^{\max} - \omega_2^{\min}}{2} \right).
 \end{aligned}$$

The final conditions for the variables $P_1^j(t)$ and $P_2^j(t)$ are get from the boundary of the safe set as

$$P_1^j(0) = \langle d_1 l(g), g_1 \xi^j \rangle, \quad P_2^j(0) = \langle d_2 l(g), g_2 \xi^j \rangle,$$

where d_1 is the derivative of l taken with respect to its first argument g_1 only (and similarly for d_2). In this way, $P_1^j(t)$ and $P_2^j(t)$ are get for $t \leq 0$. Once this has been calculated, the optimal input $u^*(t)$ and the worst disturbance $d^*(t)$ are given respectively as

$$\begin{aligned}
 u^*(t) &= \begin{cases} \omega_1^*(t) = \begin{cases} \omega_1^{\max} & \text{if } P_1^1(t) > 0 \\ \omega_1^{\min} & \text{if } P_1^1(t) < 0 \end{cases} \\ v_1^*(t) = \begin{cases} v_1^{\max} & \text{if } P_1^2(t) > 0 \\ v_1^{\min} & \text{if } P_1^2(t) < 0 \end{cases} \end{cases} \\
 d^*(t) &= \begin{cases} \omega_2^*(t) = \begin{cases} \omega_2^{\max} & \text{if } P_2^1(t) > 0 \\ \omega_2^{\min} & \text{if } P_2^1(t) < 0 \end{cases} \\ v_2^*(t) = \begin{cases} v_2^{\max} & \text{if } P_2^2(t) > 0 \\ v_2^{\min} & \text{if } P_2^2(t) < 0 \end{cases} \end{cases}.
 \end{aligned}$$

3.8.5.4 Nash Solutions for Multi-Vehicle Manoeuvres

The methodology introduced in the previous sections can be generalized to find conflict-resolutions for multi-vehicle manoeuvres. Consider the three-

vehicle dynamics:

$$\dot{g} = gX, \quad (3.64)$$

with

$$g = \begin{bmatrix} g_1 & 0 & 0 \\ 0 & g_2 & 0 \\ 0 & 0 & g_3 \end{bmatrix}, \quad X = \begin{bmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ 0 & 0 & X_3 \end{bmatrix},$$

where g is an element in the configuration space $M = SE(2) \times SE(2) \times SE(2)$ and $X \in \mathfrak{se}(2) \times \mathfrak{se}(2) \times \mathfrak{se}(2)$ is linearly parameterised by inputs (ω_1, v_1) , (ω_2, v_2) and (ω_3, v_3) .

Now, the target set T is defined as

$$T = \{g \in M \mid l_1(g) < 0 \vee l_2(g) < 0 \vee l_3(g) < 0\},$$

where

$$\begin{aligned} l_1(g) &= \min\{(x_2 - x_1)^2 + (y_2 - y_1)^2 - \rho^2, \quad (x_3 - x_1)^2 + (y_3 - y_1)^2 - \rho^2\}, \\ l_2(g) &= \min\{(x_3 - x_2)^2 + (y_3 - y_2)^2 - \rho^2, \quad (x_1 - x_2)^2 + (y_1 - y_2)^2 - \rho^2\}, \\ l_3(g) &= \min\{(x_2 - x_3)^2 + (y_2 - y_3)^2 - \rho^2, \quad (x_1 - x_3)^2 + (y_1 - y_3)^2 - \rho^2\}. \end{aligned}$$

The control inputs $u = (u_1, u_2, u_3)$ are the actions of vehicle 1, 2 and 3:

$$u_i = (\omega_i, v_i) \in U_i,$$

where U_i are defined as

$$U_i = ([\omega_i^{\min}, \omega_i^{\max}], [v_i^{\min}, v_i^{\max}]).$$

Clearly, this can be generalized to N vehicles.

The cost functions $J_i(g, \{u_i(\cdot)\}, t)$ are defined as

$$J_i(g, \{u_i(\cdot)\}, t) : \prod_{i=1}^N SE_i(2) \times \prod_{i=1}^N \mathcal{U}_i \times \mathbb{R}_+ \rightarrow \mathbb{R},$$

such that $J_i(g, \{u_i(\cdot)\}, t) = l_i(g(0))$.

The simplest non-cooperative solution strategy is a so-called *non-cooperative Nash equilibrium* (see e.g., [Basar and Olsder (1995)]). A set of controls u_i^* , ($i = 1, \dots, N$) is said to be a *Nash strategy*, if for each player modification of that strategy under the assumption that the others

play their Nash strategies results in a decrease in his payoff, that is for $i = 1, \dots, N$, and $\forall u_i(\cdot)$,

$$J_i(u_1, \dots, u_i, \dots, u_N) \leq J_i(u_1^*, \dots, u_i^*, \dots, u_N^*), \quad (u \neq u^*).$$

(Note that Nash equilibria may not be unique. It is also easy to see that for the two-player zero-sum game, a Nash equilibrium is a saddle solution with $J = J_1 = -J_2$.)

For N vehicles, the momentum functions are defined as in the two-vehicle case:

$$P_i^j = \langle p_i, g_i \xi^j \rangle,$$

with $p_i \in \mathfrak{se}(2)^*$ for $i = 1, \dots, N$ and ξ^j defined as above.

Then the Hamiltonian $H(g, p, u_1, \dots, u_N)$ is given by

$$H(g, p, u_1, \dots, u_N) = P_i^1 \omega_i + P_i^2 v_i.$$

The first case we consider is one in which all the vehicles are *cooperating*, meaning that each tries to avoid conflict assuming the others are doing the same. In this case, the optimal Hamiltonian $H^*(g, p)$ is

$$H^*(g, p) = \max_{u_i \in U_i} H(g, p, u_1, \dots, u_N).$$

For example, if $N = 3$, one may solve for (u_1^*, u_2^*, u_3^*) , on the 9D quotient space T^*M/M , so that the optimal control inputs are given as

$$u_i^*(t) = \begin{cases} \omega_i^*(t) = \begin{cases} \omega_i^{\max} & \text{if } P_i^1(t) > 0 \\ \omega_i^{\min} & \text{if } P_i^1(t) < 0 \end{cases} \\ v_i^*(t) = \begin{cases} v_i^{\max} & \text{if } P_i^2(t) > 0 \\ v_i^{\min} & \text{if } P_i^2(t) < 0 \end{cases} \end{cases}.$$

One possibility for the optimal Hamiltonian corresponding to the *non-cooperative case* is

$$H^*(g, p) = \max_{u_1 \in U_1} \max_{u_2 \in U_2} \max_{u_3 \in U_3} H(g, p, u_1, u_2, u_3).$$

3.8.6 Classical Lie Theory

In this section we present the basics of *classical theory of Lie groups* and their Lie algebras, as developed mainly by Sophus Lie, Elie Cartan, Felix Klein, Wilhelm Killing and Hermann Weyl. For more comprehensive treatment see e.g., [Chevalley (1955); Helgason (2001)].

3.8.6.1 Basic Tables of Lie Groups and Their Lie Algebras

One classifies Lie groups regarding their algebraic properties (simple, semisimple, solvable, nilpotent, Abelian), their connectedness (connected or simply connected) and their compactness (see Tables A.1–A.3). This is the content of the *Hilbert 5th problem* (see, e.g., [Weisstein (2004); Wikipedia (2005)]).

Some real Lie groups and their Lie algebras:

Lie group	Description	Remarks	Lie algb.	Description	dim / \mathbb{R}
\mathbb{R}^n	Euclidean space with addition	Abelian, simply connected, not compact	\mathbb{R}^n	the Lie bracket is zero	n
\mathbb{R}^\times	nonzero real numbers with multiplication	Abelian, not connected, not compact	\mathbb{R}	the Lie bracket is zero	1
$\mathbb{R}^{>0}$	positive real numbers with multiplication	Abelian, simply connected, not compact	\mathbb{R}	the Lie bracket is zero	1
$S^1 = \mathbb{R}/\mathbb{Z}$	complex numbers of absolute value 1, with multiplication	Abelian, connected, not simply connected, compact	\mathbb{R}	the Lie bracket is zero	1
\mathbb{H}^\times	non-zero quaternions with multiplication	simply connected, not compact	\mathbb{H}	quaternions, with Lie bracket the commutator	4
S^3	quaternions of absolute value 1, with multiplication; a 3-sphere	simply connected, compact, simple and semi-simple, isomorphic to $SU(2)$, $SO(3)$ and to $Spin(3)$	\mathbb{R}^3	real 3-vectors, with Lie bracket the cross product; isomorphic to $\mathfrak{su}(2)$ and to $\mathfrak{so}(3)$	3
$GL(n, \mathbb{R})$	general linear group: invertible n -by- n real matrices	not connected, not compact	$M(n, \mathbb{R})$	n -by- n matrices, with Lie bracket the commutator	n^2
$GL^+(n, \mathbb{R})$	n -by- n real matrices with positive determinant	simply connected, not compact	$M(n, \mathbb{R})$	n -by- n matrices, with Lie bracket the commutator	n^2

Classical real Lie groups and their Lie algebras:

Lie group	Description	Remarks	Lie algb.	Description	dim / \mathbb{R}
$SL(n, \mathbb{R})$	special linear group: real matrices with determinant 1	simply connected, not compact if $n > 1$	$\mathfrak{sl}(n, \mathbb{R})$	square matrices with trace 0, with Lie bracket the commutator	$n^2 - 1$
$O(n, \mathbb{R})$	orthogonal group: real orthogonal matrices	not connected, compact	$\mathfrak{so}(n, \mathbb{R})$	skew-symmetric square real matrices, with Lie bracket the commutator; $\mathfrak{so}(3, \mathbb{R})$ is isomorphic to $\mathfrak{su}(2)$ and to \mathbb{R}^3 with the cross product	$n(n-1)/2$
$SO(n, \mathbb{R})$	special orthogonal group: real orthogonal matrices with determinant 1	connected, compact, for $n \geq 2$: not simply connected, for $n = 3$ and $n \geq 5$: simple and semisimple	$\mathfrak{so}(n, \mathbb{R})$	skew-symmetric square real matrices, with Lie bracket the commutator	$n(n-1)/2$
$Spin(n)$	spinor group	simply connected, compact, for $n = 3$ and $n \geq 5$: simple and semisimple	$\mathfrak{so}(n, \mathbb{R})$	skew-symmetric square real matrices, with Lie bracket the commutator	$n(n-1)/2$
$U(n)$	unitary group: complex unitary n -by- n matrices	isomorphic to S^1 for $n = 1$, not simply connected, compact	$\mathfrak{u}(n)$	square complex matrices A satisfying $A = -A^*$, with Lie bracket the commutator	n^2
$SU(n)$	special unitary group: complex unitary n -by- n matrices with determinant 1	simply connected, compact, for $n \geq 2$: simple and semisimple	$\mathfrak{su}(n)$	square complex matrices A with trace 0 satisfying $A = -A^*$, with Lie bracket the commutator	$n^2 - 1$

Basic complex Lie groups and their Lie algebras:⁶

Lie group	Description	Remarks	Lie algb.	Description	dim / \mathbb{C}
\mathbb{C}^n	group operation is addition	Abelian, simply connected, not compact	\mathbb{C}^n	the Lie bracket is zero	n
\mathbb{C}^\times	nonzero complex numbers with multiplication	Abelian, not simply connected, not compact	\mathbb{C}	the Lie bracket is zero	1
$GL(n, \mathbb{C})$	general linear group: invertible n -by- n complex matrices	simply connected, not compact, for $n = 1$: isomorphic to \mathbb{C}^\times	$M(n, \mathbb{C})$	n -by- n matrices, with Lie bracket the commutator	n^2
$SL(n, \mathbb{C})$	special linear group: complex matrices with determinant 1	simple, semisimple, simply connected, for $n \geq 2$: not compact	$\mathfrak{sl}(n, \mathbb{C})$	square matrices with trace 0, with Lie bracket the commutator	$n^2 - 1$
$O(n, \mathbb{C})$	orthogonal group: complex orthogonal matrices	not connected, for $n \geq 2$: not compact	$\mathfrak{so}(n, \mathbb{C})$	skew-symmetric square complex matrices, with Lie bracket the commutator	$n(n-1)/2$
$SO(n, \mathbb{C})$	special orthogonal group: complex orthogonal matrices with determinant 1	for $n \geq 2$: not compact, not simply connected, for $n = 3$ and $n \geq 5$: simple and semisimple	$\mathfrak{so}(n, \mathbb{C})$	skew-symmetric square complex matrices, with Lie bracket the commutator	$n(n-1)/2$

3.8.6.2 Representations of Lie groups

The idea of a *representation of a Lie group* plays an important role in the study of continuous symmetry (see, e.g., [Helgason (2001)]). A great deal is known about such representations, a basic tool in their study being the use of the corresponding 'infinitesimal' representations of Lie algebras.

Formally, a representation of a Lie group G on a vector space V (over a field K) is a group homomorphism $G \rightarrow \text{Aut}(V)$ from G to the automorphism group of V . If a basis for the vector space V is chosen, the representation can be expressed as a homomorphism into $GL(n, K)$. This is known as a *matrix representation*.

⁶The dimensions given are dimensions over \mathbb{C} . Note that every complex Lie group/algebra can also be viewed as a real Lie group/algebra of twice the dimension.

On the Lie algebra level, there is a corresponding linear map from the Lie algebra of G to $\text{End}(V)$ preserving the Lie bracket $[\cdot, \cdot]$.

If the homomorphism is in fact an monomorphism, the representation is said to be *faithful*.

A unitary representation is defined in the same way, except that G maps to unitary matrices; the Lie algebra will then map to skew-Hermitian matrices.

Now, if G is a semisimple group, its finite-dimensional representations can be decomposed as direct sums of irreducible representations. The irreducibles are indexed by highest weight; the allowable (*dominant*) highest weights satisfy a suitable positivity condition. In particular, there exists a set of *fundamental weights*, indexed by the vertices of the *Dynkin diagram* of G (see below), such that dominant weights are simply non-negative integer linear combinations of the fundamental weights.

If G is a commutative compact Lie group, then its irreducible representations are simply the continuous characters of G . A *quotient representation* is a quotient module of the group ring.

3.8.6.3 Root Systems and Dynkin Diagrams

A *root system* is a special configuration in Euclidean space that has turned out to be fundamental in Lie theory as well as in its applications. Also, the classification scheme for root systems, by *Dynkin diagrams*, occurs in parts of mathematics with no overt connection to Lie groups (such as singularity theory, see e.g., [Helgason (2001); Weisstein (2004); Wikipedia (2005)]).

Definitions

Formally, a *root system* is a finite set Φ of non-zero vectors (*roots*) spanning a finite-dimensional Euclidean space V and satisfying the following properties:

- (1) The only scalar multiples of a root α in V which belong to Φ are α itself and $-\alpha$.
- (2) For every root α in V , the set Φ is symmetric under reflection through the hyperplane of vectors perpendicular to α .
- (3) If α and β are vectors in Φ , the projection of 2β onto the line through α is an integer multiple of α .

The *rank* of a root system Φ is the dimension of V . Two root systems may be combined by regarding the Euclidean spaces they span as mutually

orthogonal subspaces of a common Euclidean space. A root system which does not arise from such a combination, such as the systems A_2 , B_2 , and G_2 in Figure 3.9, is said to be *irreducible*.

Two irreducible root systems (V_1, Φ_1) and (V_2, Φ_2) are considered to be the same if there is an invertible linear transformation $V_1 \rightarrow V_2$ which preserves distance up to a scale factor and which sends Φ_1 to Φ_2 .

The group of isometries of V generated by reflections through hyperplanes associated to the roots of Φ is called the Weyl group of Φ as it acts faithfully on the finite set Φ , the Weyl group is always finite.

Classification

It is not too difficult to classify the root systems of rank 2 (see Figure 3.9).

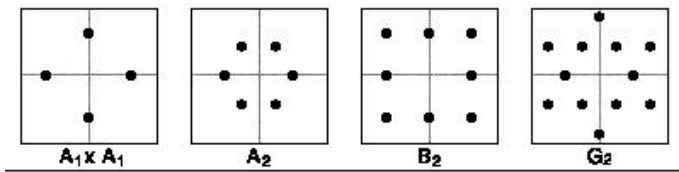


Fig. 3.9 Classification of root systems of rank 2.

Whenever Φ is a root system in V and W is a subspace of V spanned by $\Psi = \Phi \cap W$, then Ψ is a root system in W . Thus, our exhaustive list of root systems of rank 2 shows the geometric possibilities for any two roots in a root system. In particular, two such roots meet at an angle of 0, 30, 45, 60, 90, 120, 135, 150, or 180 degrees.

In general, irreducible root systems are specified by a family (indicated by a letter A to G) and the rank (indicated by a subscript n). There are four *infinite families*:

- A_n ($n \geq 1$), which corresponds to the special unitary group, $SU(n+1)$;
- B_n ($n \geq 2$), which corresponds to the special orthogonal group, $SO(2n+1)$;
- C_n ($n \geq 3$), which corresponds to the symplectic group, $Sp(2n)$;
- D_n ($n \geq 4$), which corresponds to the special orthogonal group, $SO(2n)$,

as well as five *exceptional cases*: E_6, E_7, E_8, F_4, G_2 .

Dynkin Diagrams

A Dynkin diagram is a graph with a few different kinds of possible edges (see Figure 3.10). The connected components of the graph correspond to the irreducible subalgebras of \mathfrak{g} . So a simple Lie algebra's Dynkin diagram has only one component. The rules are restrictive. In fact, there are only certain possibilities for each component, corresponding to the classification of semi-simple Lie algebras (see, e.g., [Conway *et al.* (1985)]).

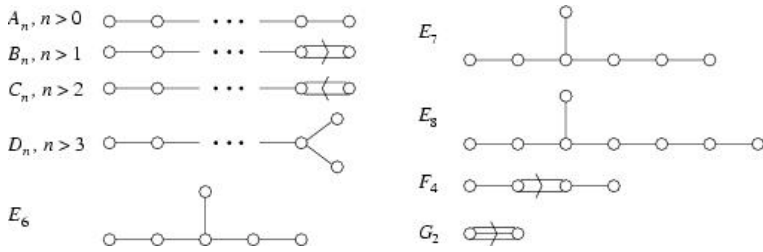


Fig. 3.10 The problem of classifying irreducible root systems reduces to the problem of classifying connected Dynkin diagrams.

The *roots* of a complex Lie algebra form a lattice of rank k in a *Cartan subalgebra* $\mathfrak{h} \subset \mathfrak{g}$, where k is the Lie algebra rank of \mathfrak{g} . Hence, the *root lattice* can be considered a lattice in \mathbb{R}^k . A vertex, or node, in the Dynkin diagram is drawn for each *Lie algebra simple root*, which corresponds to a generator of the root lattice. Between two nodes α and β , an edge is drawn if the simple roots are not perpendicular. One line is drawn if the angle between them is $2\pi/3$, two lines if the angle is $3\pi/4$, and three lines are drawn if the angle is $5\pi/6$. There are no other possible angles between Lie algebra simple roots. Alternatively, the number of lines N between the simple roots α and β is given by

$$N = A_{\alpha\beta}A_{\beta\alpha} = \frac{2\langle\alpha,\beta\rangle}{|\alpha|^2} \frac{2\langle\beta,\alpha\rangle}{|\beta|^2} = 4\cos^2\theta,$$

where $A_{\alpha\beta} = \frac{2\langle\alpha,\beta\rangle}{|\alpha|^2}$ is an entry in the *Cartan matrix* ($A_{\alpha\beta}$) (for details on Cartan matrix see, e.g., [Helgason (2001); Weisstein (2004)]). In a Dynkin diagram, an arrow is drawn from the longer root to the shorter root (when the angle is $3\pi/4$ or $5\pi/6$).

Here are some properties of *admissible Dynkin diagrams*:

- (1) A diagram obtained by removing a node from an admissible diagram

is admissible.

- (2) An admissible diagram has no loops.
- (3) No node has more than three lines attached to it.
- (4) A sequence of nodes with only two single lines can be collapsed to give an admissible diagram.
- (5) The only connected diagram with a triple line has two nodes.

A *Coxeter–Dynkin diagram*, also called a *Coxeter graph*, is the same as a Dynkin diagram, but without the arrows. The Coxeter diagram is sufficient to characterize the algebra, as can be seen by enumerating connected diagrams.

The simplest way to recover a simple Lie algebra from its Dynkin diagram is to first reconstruct its Cartan matrix (A_{ij}) . The i th node and j th node are connected by $A_{ij}A_{ji}$ lines. Since $A_{ij} = 0$ iff $A_{ji} = 0$, and otherwise $A_{ij} \in \{-3, -2, -1\}$, it is easy to find A_{ij} and A_{ji} , up to order, from their product. The arrow in the diagram indicates which is larger. For example, if node 1 and node 2 have two lines between them, from node 1 to node 2, then $A_{12} = -1$ and $A_{21} = -2$.

However, it is worth pointing out that each simple Lie algebra can be constructed concretely. For instance, the infinite families A_n , B_n , C_n , and D_n correspond to the special linear Lie algebra $\mathfrak{gl}(n+1, \mathbb{C})$, the odd orthogonal Lie algebra $\mathfrak{so}(2n+1, \mathbb{C})$, the symplectic Lie algebra $\mathfrak{sp}(2n, \mathbb{C})$, and the even orthogonal Lie algebra $\mathfrak{so}(2n, \mathbb{C})$. The other simple Lie algebras are called *exceptional Lie algebras*, and have constructions related to the *octonions*.

To prove this classification Theorem, one uses the angles between pairs of roots to encode the root system in a much simpler combinatorial object, the Dynkin diagram. The Dynkin diagrams can then be classified according to the scheme given above.

To every root system is associated a corresponding Dynkin diagram. Otherwise, the Dynkin diagram can be extracted from the root system by choosing a *base*, that is a subset Δ of Φ which is a basis of V with the special property that every vector in Φ when written in the basis Δ has either all coefficients ≥ 0 or else all ≤ 0 .

The vertices of the Dynkin diagram correspond to vectors in Δ . An edge is drawn between each non-orthogonal pair of vectors; it is a double edge if they make an angle of 135 degrees, and a triple edge if they make an angle of 150 degrees. In addition, double and triple edges are marked with an angle sign pointing toward the shorter vector.

Although a given root system has more than one base, the Weyl group acts transitively on the set of bases. Therefore, the root system determines the Dynkin diagram. Given two root systems with the same Dynkin diagram, we can match up roots, starting with the roots in the base, and show that the systems are in fact the same.

Thus the problem of classifying root systems reduces to the problem of classifying possible Dynkin diagrams, and the problem of classifying irreducible root systems reduces to the problem of classifying connected Dynkin diagrams. Dynkin diagrams encode the inner product on E in terms of the basis Δ , and the condition that this inner product must be positive definite turns out to be all that is needed to get the desired classification (see Figure 3.10).

In detail, the individual root systems can be realized case-by-case, as in the following paragraphs:

A_n. Let V be the subspace of \mathbb{R}^{n+1} for which the coordinates sum to 0, and let Φ be the set of vectors in V of length $\sqrt{2}$ and with integer coordinates in \mathbb{R}^{n+1} . Such a vector must have all but two coordinates equal to 0, one coordinate equal to 1, and one equal to -1, so there are $n^2 + n$ roots in all.

B_n. Let $V = \mathbb{R}^n$, and let Φ consist of all integer vectors in V of length 1 or $\sqrt{2}$. The total number of roots is $2n^2$.

C_n. Let $V = \mathbb{R}^n$, and let Φ consist of all integer vectors in V of $\sqrt{2}$ together with all vectors of the form 2λ , where λ is an integer vector of length 1. The total number of roots is $2n^2$. The total number of roots is $2n^2$.

D_n. Let $V = \mathbb{R}^n$, and let Φ consist of all integer vectors in V of length $\sqrt{2}$. The total number of roots is $2n(n-1)$.

E_n. For V_8 , let $V = \mathbb{R}^8$, and let E_8 denote the set of vectors α of length $\sqrt{2}$ such that the coordinates of 2α are all integers and are either all even or all odd. Then E_7 can be constructed as the intersection of E_8 with the hyperplane of vectors perpendicular to a fixed root α in E_8 , and E_6 can be constructed as the intersection of E_8 with two such hyperplanes corresponding to roots α and β which are neither orthogonal to one another nor scalar multiples of one another. The root systems E_6 , E_7 , and E_8 have 72, 126, and 240 roots respectively.

F₄. For F_4 , let $V = \mathbb{R}^4$, and let Φ denote the set of vectors α of length 1 or $\sqrt{2}$ such that the coordinates of 2α are all integers and are either all even or all odd. There are 48 roots in this system.

G₂. There are 12 roots in G_2 , which form the vertices of a *hexagram*.

Root Systems and Lie Theory

Irreducible root systems classify a number of related objects in Lie theory, notably:

- (1) Simple complex Lie algebras;
- (2) Simple complex Lie groups;
- (3) Simply connected complex Lie groups which are simple modulo centers;
and
- (4) Simple compact Lie groups.

In each case, the roots are non-zero weights of the adjoint representation.

A root system can also be said to describe a *plant's root* and associated systems.

3.8.6.4 Simple and Semisimple Lie Groups and Algebras

A *simple Lie group* is a Lie group which is also a simple group. These groups, and groups closely related to them, include many of the so-called *classical groups* of geometry, which lie behind projective geometry and other geometries derived from it by the *Erlangen programme* of Felix Klein. They also include some *exceptional groups*, that were first discovered by those pursuing the classification of simple Lie groups. The exceptional groups account for many special examples and configurations in other branches of mathematics. In particular the classification of finite simple groups depended on a thorough prior knowledge of the 'exceptional' possibilities.

The complete listing of the simple Lie groups is the basis for the theory of the semisimple Lie groups and reductive groups, and their representation theory. This has turned out not only to be a major extension of the theory of compact Lie groups (and their representation theory), but to be of basic significance in mathematical physics.

Such groups are classified using the prior classification of the complex simple Lie algebras. It has been shown that a simple Lie group has a simple Lie algebra that will occur on the list given there, once it is complexified (that is, made into a complex vector space rather than a real one). This reduces the classification to two further matters.

The groups $SO(p, q, \mathbb{R})$ and $SO(p + q, \mathbb{R})$, for example, give rise to different real Lie algebras, but having the same Dynkin diagram. In general there may be different *real forms* of the same complex Lie algebra.

Secondly, the Lie algebra only determines uniquely the simply connected

(universal) cover G^* of the component containing the identity of a Lie group G . It may well happen that G^* is not actually a simple group, for example having a non-trivial center. We have therefore to worry about the global topology, by computing the fundamental group of G (an Abelian group: a Lie group is an H -space). This was done by Elie Cartan.

For an example, take the special orthogonal groups in even dimension. With $-I$ a scalar matrix in the center, these are not actually simple groups; and having a two-fold spin cover, they aren't simply-connected either. They lie 'between' G^* and G , in the notation above.

Recall that a *semisimple module* is a module in which each submodule is a direct summand. In particular, a *semisimple representation* is completely reducible, i.e., is a direct sum of irreducible representations (under a descending chain condition). Similarly, one speaks of an Abelian category as being semisimple when every object has the corresponding property. Also, a semisimple ring is one that is semisimple as a module over itself.

A *semisimple matrix* is diagonalizable over any algebraically closed field containing its entries. In practice this means that it has a diagonal matrix as its Jordan normal form.

A Lie algebra \mathfrak{g} is called *semisimple* when it is a direct sum of *simple Lie algebras*, i.e., non-trivial Lie algebras \mathfrak{L} whose only ideals are $\{0\}$ and \mathfrak{L} itself. An equivalent condition is that the *Killing form*

$$\mathcal{B}(X, Y) = \text{Tr}(Ad(X) Ad(Y))$$

is non-degenerate [Schafer (1996)]. The following properties can be proved equivalent for a finite-dimensional algebra \mathfrak{L} over a field of characteristic 0:

1. \mathfrak{L} is semisimple.
2. \mathfrak{L} has no nonzero Abelian ideal.
3. \mathfrak{L} has zero radical (the radical is the biggest solvable ideal).
4. Every representation of \mathfrak{L} is fully reducible, i.e., is a sum of irreducible representations.
5. \mathfrak{L} is a (finite) direct product of simple Lie algebras (a Lie algebra is called simple if it is not Abelian and has no nonzero ideal).

A *connected Lie group* is called *semisimple* when its Lie algebra is semisimple; and the same holds for algebraic groups. Every finite dimensional representation of a semisimple Lie algebra, Lie group, or algebraic group in characteristic 0 is semisimple, i.e., completely reducible, but the converse is not true. Moreover, in characteristic $p > 0$, semisimple Lie

groups and Lie algebras have finite dimensional representations which are not semisimple. An element of a semisimple Lie group or Lie algebra is itself semisimple if its image in every finite-dimensional representation is semisimple in the sense of matrices.

Every semisimple Lie algebra \mathfrak{g} can be classified by its Dynkin diagram [Helgason (2001)].

3.9 Lie Symmetries and Prolongations on Manifolds

In this section we continue our expose on Lie groups of symmetry, as a link to modern jet machinery, developed below.

3.9.1 Lie Symmetry Groups

3.9.1.1 Exponentiation of Vector Fields on M

Let $x = (x^1, \dots, x^r)$ be local coordinates at a point m on a smooth n -manifold M . Recall that the flow generated by the vector-field

$$v = \xi^i(x) \partial_{x^i} \in M,$$

is a solution of the system of ODEs

$$\frac{dx^i}{d\varepsilon} = \xi^i(x^1, \dots, x^m), \quad (i = 1, \dots, r).$$

The computation of the flow, or one-parameter group of diffeomorphisms, generated by a given vector-field v (i.e., solving the system of ODEs) is often referred to as *exponentiation of a vector-field*, denoted by $\exp(\varepsilon v)x$ (see [Olver (1986)]).

If $v, w \in M$ are two vectors defined by

$$v = \xi^i(x) \partial_{x^i} \quad \text{and} \quad w = \eta^i(x) \partial_{x^i},$$

then

$$\exp(\varepsilon v) \exp(\theta w) x = \exp(\theta w) \exp(\varepsilon v) x,$$

for all $\varepsilon, \theta \in \mathbb{R}, x \in M$, such that both sides are defined, iff they commute, i.e., $[v, w] = 0$ everywhere [Olver (1986)].

A system of vector-fields $\{v_1, \dots, v_r\}$ on a smooth manifold M is in *involution* if there exist smooth real-valued functions $h_{ij}^k(x)$, $x \in M$,

$i, j, k = 1, \dots, r$, such that for each i, j ,

$$[v_i, v_j] = h_{ij}^k \cdot v_k.$$

Let $v \neq 0$ be a right-invariant vector-field on a Lie group G . Then the flow generated by v through the identity e , namely

$$g_\varepsilon = \exp(\varepsilon v) e \equiv \exp(\varepsilon v),$$

is defined for all $\varepsilon \in \mathbb{R}$ and forms a one-parameter subgroup of G , with

$$g_{\varepsilon+\delta} = g_\varepsilon \cdot g_\delta, \quad g_0 = e, \quad g_\varepsilon^{-1} = g_{-\varepsilon},$$

isomorphic to either \mathbb{R} itself or the circle group $SO(2)$. Conversely, any connected 1D subgroup of G is generated by such a right-invariant vector-field in the above manner [Olver (1986)].

For example, let $G = GL(n)$ with Lie algebra $\mathfrak{gl}(n)$, the space of all $n \times n$ matrices with commutator as the Lie bracket. If $A \in \mathfrak{gl}(n)$, then the corresponding right-invariant vector-field v_A on $GL(n)$ has the expression [Olver (1986)]

$$v_A = a_k^i x_j^k \partial_{x_j^i}.$$

The one-parameter subgroup $\exp(\varepsilon v_A) e$ is found by integrating the system of n^2 ordinary differential equations

$$\frac{dx_j^i}{d\varepsilon} = a_k^i x_j^k, \quad x_j^i(0) = \delta_j^i, \quad (i, j = 1, \dots, n),$$

involving matrix entries of A . The solution is just the matrix exponential $X(\varepsilon) = e^{\varepsilon A}$, which is the one-parameter subgroup of $GL(n)$ generated by a matrix A in $\mathfrak{gl}(n)$.

Recall that the *exponential map* $\exp : \mathfrak{g} \rightarrow G$ is get by setting $\varepsilon = 1$ in the one-parameter subgroup generated by vector-field v :

$$\exp(v) \equiv \exp(v) e.$$

Its differential at 0,

$$d\exp : T\mathfrak{g}|_0 \simeq \mathfrak{g} \rightarrow TG|_e \simeq \mathfrak{g}$$

is the identity map.

3.9.1.2 Lie Symmetry Groups and General DEs

Consider a system \mathcal{S} of general differential equations (DEs, to be distinguished from ODEs) involving p independent variables $x = (x^1, \dots, x^p)$, and q dependent variables $u = (u^1, \dots, u^q)$. The solution of the system will be of the form $u = f(x)$, or, in components, $u^\alpha = f^\alpha(x^1, \dots, x^p)$, $\alpha = 1, \dots, q$ (so that Latin indices refer to independent variables while Greek indices refer to dependent variables). Let $X = \mathbb{R}^p$, with coordinates $x = (x^1, \dots, x^p)$, be the space representing the independent variables, and let $U = \mathbb{R}^q$, with coordinates $u = (u^1, \dots, u^q)$, represent dependent variables. A Lie symmetry group G of the system \mathcal{S} will be a local group of transformations acting on some open subset $M \subset X \times U$ in such way that G transforms solutions of \mathcal{S} to other solutions of \mathcal{S} [Olver (1986)].

More precisely, we need to explain exactly how a given transformation $g \in G$, where G is a Lie group, transforms a function $u = f(x)$. We firstly identify the function $u = f(x)$ with its graph

$$\Gamma_f \equiv \{(x, f(x)) : x \in \text{dom } f \equiv \Omega\} \subset X \times U,$$

where Γ_f is a submanifold of $X \times U$. If $\Gamma_f \subset M_g \equiv \text{dom } g$, then the transform of Γ_f by g is defined as

$$g \cdot \Gamma_f = \{(\tilde{x}, \tilde{u}) = g \cdot (x, u) : (x, u) \in \Gamma_f\}.$$

We write $\tilde{f} = g \cdot f$ and call the function \tilde{f} the *transform* of f by g .

For example, let $p = 1$ and $q = 1$, so $X = \mathbb{R}$ with a single independent variable x , and $U = \mathbb{R}$ with a single dependent variable u , so we have a single ODE involving a single function $u = f(x)$. Let $G = SO(2)$ be the rotation group acting on $X \times U \simeq \mathbb{R}^2$. The transformations in G are given by

$$(\tilde{x}, \tilde{u}) = \theta \cdot (x, u) = (x \cos \theta - u \sin \theta, x \sin \theta + u \cos \theta).$$

Let $u = f(x)$ be a function whose graph is a subset $\Gamma_f \subset X \times U$. The group $SO(2)$ acts on f by rotating its graph.

In general, the procedure for finding the transformed function $\tilde{f} = g \cdot f$ is given by [Olver (1986)]:

$$g \cdot f = [\Phi_g \circ (1 \times f)] \circ [\Xi_g \circ (1 \times f)]^{-1}, \quad (3.65)$$

where $\Xi_g = \Xi_g(x, u)$, $\Phi_g = \Phi_g(x, u)$ are smooth functions such that

$$(\tilde{x}, \tilde{u}) = g \cdot (x, u) = (\Xi_g(x, u), \Phi_g(x, u)),$$

while 1 denotes the identity function of X , so $1(x) = x$. Formula (3.65) holds whenever the second factor is invertible.

Let \mathcal{S} be a system of DEs. A *symmetry group* of the system \mathcal{S} is a local Lie group of transformations G acting on an open subset $M \subset X \times U$ of the space $X \times U$ of independent and dependent variables of the system with the property that whenever $u = f(x)$ is a solution of \mathcal{S} , and whenever $g \cdot f$ is defined for $g \in G$, then $u = g \cdot f(x)$ is also a solution of the system.

For example, in the case of the ODE $u_{xx} = 0$, the rotation group $SO(2)$ is obviously a symmetry group, since the solutions are all linear functions and $SO(2)$ takes any linear function to another linear function. Another easy example is given by the classical *heat equation* $u_t = u_{xx}$. Here the group of translations

$$(x, t, u) \mapsto (x + \varepsilon a, t + \varepsilon b, u), \quad \varepsilon \in \mathbb{R},$$

is a symmetry group since $u = f(x - \varepsilon a, t - \varepsilon b)$ is a solution to the heat equation whenever $u = f(x, t)$ is.

3.9.2 Prolongations

3.9.2.1 Prolongations of Functions

Given a smooth real-valued function $u = f(x) = f(x^1, \dots, x^p)$ of p independent variables, there is an induced function $u^{(n)} = \mathbf{pr}^{(n)}f(x)$, called the *n*th *prolongation* of f [Olver (1986)], which is defined by the equations

$$u_J = \partial_J f(x) = \frac{\partial^k f(x)}{\partial x^{j_1} \partial x^{j_2} \dots \partial x^{j_k}},$$

where the *multi-index* $J = (j_1, \dots, j_k)$ is an unordered k -tuple of integers, with entries $1 \leq j_k \leq p$ indicating which derivatives are being taken. More generally, if $f : X \rightarrow U$ is a smooth function from $X \simeq \mathbb{R}^p$ to $U \simeq \mathbb{R}^q$, so $u = f(x) = f(f^1(x), \dots, f^q(x))$, there are $q \cdot p_k$ numbers

$$u_J^\alpha = \partial_J f^\alpha(x) = \frac{\partial^k f^\alpha(x)}{\partial x^{j_1} \partial x^{j_2} \dots \partial x^{j_k}},$$

needed to represent all the different k th order derivatives of the components of f at a point x . Thus $\mathbf{pr}^{(n)}f : X \rightarrow U^{(n)}$ is a function from X to the space $U^{(n)}$, and for each $x \in X$, $\mathbf{pr}^{(n)}f(x)$ is a vector whose $q \cdot p^{(n)}$ entries represent the values of f and all its derivatives up to order n at the point x .

For example, in the case $p = 2$, $q = 1$ we have $X \simeq \mathbb{R}^2$ with coordinates $(x^1, x^2) = (x, y)$, and $U \simeq \mathbb{R}$ with the single coordinate $u = f(x, y)$. The

second prolongation $u^{(2)} = \mathbf{pr}^{(2)}f(x, y)$ is given by [Olver (1986)]

$$(u; u_x, u_y; u_{xx}, u_{xy}, u_{yy}) = \left(f; \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}; \frac{\partial^2 f}{\partial x^2}, \frac{\partial^2 f}{\partial x \partial y}, \frac{\partial^2 f}{\partial y^2} \right), \quad (3.66)$$

all evaluated at (x, y) .

The n th prolongation $\mathbf{pr}^{(n)}f(x)$ is also known as the n -jet of f . In other words, the n th prolongation $\mathbf{pr}^{(n)}f(x)$ represents the Taylor polynomial of degree n for f at the point x , since the derivatives of order $\leq n$ determine the Taylor polynomial and vice versa.

3.9.2.2 Prolongations of Differential Equations

A system \mathcal{S} of n th order DEs in p independent and q dependent variables is given as a system of equations [Olver (1986)]

$$\Delta_r(x, u^{(n)}) = 0, \quad (r = 1, \dots, l), \quad (3.67)$$

involving $x = (x^1, \dots, x^p)$, $u = (u^1, \dots, u^q)$ and the derivatives of u with respect to x up to order n . The functions $\Delta(x, u^{(n)}) = (\Delta_1(x, u^{(n)}), \dots, \Delta_l(x, u^{(n)}))$ are assumed to be smooth in their arguments, so $\Delta : X \times U^{(n)} \rightarrow \mathbb{R}^l$ represents a smooth map from the *jet space* $X \times U^{(n)}$ to some l D Euclidean space (see section 4.14.12.5 below). The DEs themselves tell where the given map Δ vanishes on the jet space $X \times U^{(n)}$, and thus determine a submanifold

$$\mathcal{S}_\Delta = \left\{ (x, u^{(n)}) : \Delta(x, u^{(n)}) = 0 \right\} \subset X \times U^{(n)} \quad (3.68)$$

of the total the jet space $X \times U^{(n)}$.

We can identify the system of DEs (3.67) with its corresponding submanifold \mathcal{S}_Δ (3.68). From this point of view, a smooth *solution* of the given system of DEs is a smooth function $u = f(x)$ such that [Olver (1986)]

$$\Delta_r(x, \mathbf{pr}^{(n)}f(x)) = 0, \quad (r = 1, \dots, l),$$

whenever x lies in the domain of f . This is just a restatement of the fact that the derivatives $\partial_J f^\alpha(x)$ of f must satisfy the algebraic constraints imposed by the system of DEs. This condition is equivalent to the statement that the graph of the prolongation $\mathbf{pr}^{(n)}f(x)$ must lie entirely within the submanifold \mathcal{S}_Δ determined by the system:

$$\Gamma_f^{(n)} \equiv \left\{ (x, \mathbf{pr}^{(n)}f(x)) \right\} \subset \mathcal{S}_\Delta = \left\{ \Delta(x, u^{(n)}) = 0 \right\}.$$

We can thus take an n th order system of DEs to be a submanifold \mathcal{S}_Δ in the n -jet space $X \times U^{(n)}$ and a solution to be a function $u = f(x)$ such that the graph of the n th prolongation $\mathbf{pr}^{(n)}f(x)$ is contained in the submanifold \mathcal{S}_Δ .

For example, consider the case of *Laplace equation* in the plane

$$u_{xx} + u_{yy} = 0 \quad (\text{remember, } u_x \equiv \partial_x u).$$

Here $p = 2$ since there are two independent variables x and y , and $q = 1$ since there is one dependent variable u . Also $n = 2$ since the equation is second-order, so $\mathcal{S}_\Delta \subset X \times U^{(2)}$ is given by (3.66). A solution $u = f(x, y)$ must satisfy

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0$$

for all (x, y) . This is the same as requiring that the graph of the second prolongation $\mathbf{pr}^{(2)}f$ lie in \mathcal{S}_Δ .

3.9.2.3 Prolongations of Group Actions

Let G be a local group of transformations acting on an open subset $M \subset X \times U$ of the space of independent and dependent variables. There is an induced local action of G on the n -jet space $M^{(n)}$, called the n th prolongation $\mathbf{pr}^{(n)}G$ of the action of G on M . This prolongation is defined so that it transforms the derivatives of functions $u = f(x)$ into the corresponding derivatives of the transformed function $\tilde{u} = \tilde{f}(\tilde{x})$ [Olver (1986)].

More precisely, suppose $(x_0, u_0^{(n)})$ is a given point in $M^{(n)}$. Choose any smooth function $u = f(x)$ defined in a neighborhood of x_0 , whose graph Γ_f lies in M , and has the given derivatives at x_0 :

$$u_0^{(n)} = \mathbf{pr}^{(n)}f(x_0), \quad \text{i.e.,} \quad u_{j_0}^\alpha = \partial_J f^\alpha(x_0).$$

If g is an element of G sufficiently near the identity, the transformed function $g \cdot f$ as given by (3.65) is defined in a neighborhood of the corresponding point $(\tilde{x}_0, \tilde{u}_0) = g \cdot (x_0, u_0)$, with $u_0 = f(x_0)$ being the zeroth order components of $u_0^{(n)}$. We then determine the action of the prolonged group of transformations $\mathbf{pr}^{(n)}g$ on the point $(x_0, u_0^{(n)})$ by evaluating the derivatives of the transformed function $g \cdot f$ at \tilde{x}_0 ; explicitly [Olver (1986)]

$$\mathbf{pr}^{(n)}g \cdot (x_0, u_0^{(n)}) = (\tilde{x}_0, \tilde{u}_0^{(n)}),$$

where

$$\tilde{u}_0^{(n)} \equiv \mathbf{pr}^{(n)}(g \cdot f)(\tilde{x}_0).$$

For example, let $p = q = 1$, so $X \times U \simeq \mathbb{R}^2$, and consider the action of the rotation group $SO(2)$. To calculate its first prolongation $\mathbf{pr}^{(1)}SO(2)$, first note that $X \times U^{(1)} \simeq \mathbb{R}^3$, with coordinates (x, u, u_x) . Given a function $u = f(x)$, the first prolongation is [Olver (1986)]

$$\mathbf{pr}^{(1)}f(x) = (f(x), f'(x)).$$

Now, given a point $(x^0, u^0, u_x^0) \in X \times U^{(1)}$, and a rotation in $SO(2)$ characterized by the angle θ as given above, the corresponding transformed point

$$\mathbf{pr}^{(1)}\theta \cdot (x^0, u^0, u_x^0) = (\tilde{x}^0, \tilde{u}^0, \tilde{u}_x^0)$$

(provided it exists). As for the first-order derivative, we find

$$\tilde{u}_x^0 = \frac{\sin \theta + u_x \cos \theta}{\cos \theta - u_x \sin \theta}.$$

Now, applying the group transformations given above, and dropping the 0-indices, we find that the prolonged action $\mathbf{pr}^{(1)}SO(2)$ on $X \times U^{(1)}$ is given by

$$\mathbf{pr}^{(1)}\theta \cdot (x, u, u_x) = \left(x \cos \theta - u \sin \theta, x \sin \theta + u \cos \theta, \frac{\sin \theta + u_x \cos \theta}{\cos \theta - u_x \sin \theta} \right),$$

which is defined for $|\theta| < |\operatorname{arccot} u_x|$. Note that even though $SO(2)$ is a linear, globally defined group of transformations, its first prolongation $\mathbf{pr}^{(1)}SO(2)$ is both nonlinear and only locally defined. This fact demonstrates the complexity of the operation of prolonging a group of transformations.

In general, for any Lie group G , the first prolongation $\mathbf{pr}^{(1)}G$ acts on the original variables (x, u) exactly the same way that G itself does; only the action on the derivative u_x gives an new information. Therefore, $\mathbf{pr}^{(0)}G$ agrees with G itself, acting on $M^{(0)} = M$.

3.9.2.4 Prolongations of Vector Fields

Prolongation of the infinitesimal generators of the group action turn out to be the *infinitesimal generators* of the *prolonged group action* [Olver (1986)]. Let $M \subset X \times U$ be open and suppose v is a vector-field on M , with

corresponding local one-parameter group $\exp(\varepsilon v)$. The n th prolongation of v , denoted $\mathbf{pr}^{(n)}v$, will be a vector-field on the n -jet space $M^{(n)}$, and is defined to be the infinitesimal generator of the corresponding prolonged one-parameter group $\mathbf{pr}^{(n)}[\exp(\varepsilon v)]$. In other words,

$$\mathbf{pr}^{(n)}v|_{(x, u^{(n)})} = \left. \frac{d}{d\varepsilon} \right|_{\varepsilon=0} \mathbf{pr}^{(n)}[\exp(\varepsilon v)](x, u^{(n)}) \quad (3.69)$$

for any $(x, u^{(n)}) \in M^{(n)}$.

For a vector-field v on M , given by

$$v = \xi^i(x, u) \frac{\partial}{\partial x^i} + \phi^\alpha(x, u) \frac{\partial}{\partial u^\alpha}, \quad (i = 1, \dots, p, \alpha = 1, \dots, q),$$

the n th prolongation $\mathbf{pr}^{(n)}v$ is given by [Olver (1986)]

$$\mathbf{pr}^{(n)}v = \xi^i(x, u) \frac{\partial}{\partial x^i} + \phi_J^\alpha(x, u^{(n)}) \frac{\partial}{\partial u_J^\alpha},$$

with $\phi_0^\alpha = \phi^\alpha$, and J a multiindex defined above.

For example, in the case of $SO(2)$ group, the corresponding infinitesimal generator is

$$v = -u \frac{\partial}{\partial x} + x \frac{\partial}{\partial u},$$

with

$$\exp(\varepsilon v)(x, u) = (x \cos \varepsilon - u \sin \varepsilon, x \sin \varepsilon + u \cos \varepsilon),$$

being the rotation through angle ε . The first prolongation takes the form

$$\mathbf{pr}^{(1)}[\exp(\varepsilon v)](x, u, u_x) = \left(x \cos \varepsilon - u \sin \varepsilon, x \sin \varepsilon + u \cos \varepsilon, \frac{\sin \varepsilon + u_x \cos \varepsilon}{\cos \varepsilon - u_x \sin \varepsilon} \right).$$

According to (3.69), the first prolongation of v is get by differentiating these expressions with respect to ε and setting $\varepsilon = 0$, which gives

$$\mathbf{pr}^{(1)}v = -u \frac{\partial}{\partial x} + x \frac{\partial}{\partial u} + (1 + u_x^2) \frac{\partial}{\partial u_x}.$$

3.9.2.5 General Prolongation Formula

Let

$$v = \xi^i(x, u) \frac{\partial}{\partial x^i} + \phi^\alpha(x, u) \frac{\partial}{\partial u^\alpha}, \quad (i = 1, \dots, p, \alpha = 1, \dots, q), \quad (3.70)$$

be a vector-field defined on an open subset $M \subset X \times U$. The n th prolongation of v is the vector-field [Olver (1986)]

$$\mathbf{pr}^{(n)}v = v + \phi_J^\alpha(x, u^{(n)}) \frac{\partial}{\partial u_J^\alpha}, \quad (3.71)$$

defined on the corresponding jet space $M^{(n)} \subset X \times U^{(n)}$. The coefficient functions ϕ_J^α are given by the following formula:

$$\phi_J^\alpha = D_J (\phi^\alpha - \xi^i u_{J,i}^\alpha) + \xi^i u_{J,i}^\alpha, \quad (3.72)$$

where $u_i^\alpha = \partial u^\alpha / \partial x^i$, and $u_{J,i}^\alpha = \partial u_J^\alpha / \partial x^i$. D_J is the *total derivative* with respect to the multiindex J , i.e.,

$$D_J = D_{j_1} D_{j_2} \dots D_{j_k},$$

while the total derivative with respect to the ordinary index, D_i , is defined as follows. Let $P(x, u^{(n)})$ be a smooth function of x, u and derivatives of u up to order n , defined on an open subset $M^{(n)} \subset X \times U^{(n)}$. the total derivative of P with respect to x^i is the unique smooth function $D_i P(x, u^{(n)})$ defined on $M^{(n+1)}$ and depending on derivatives of u up to order $n+1$, with the *recursive property* that if $u = f(x)$ is any smooth function then

$$D_i P(x, \mathbf{pr}^{(n+1)}f(x)) = \partial_{x^i} \{P(x, \mathbf{pr}^{(n)}f(x))\}.$$

For example, in the case of $SO(2)$ group, with the infinitesimal generator

$$v = -u \frac{\partial}{\partial x} + x \frac{\partial}{\partial u},$$

the first prolongation is (as calculated above)

$$\mathbf{pr}^{(1)}v = -u \frac{\partial}{\partial x} + x \frac{\partial}{\partial u} + \phi^x \frac{\partial}{\partial u_x},$$

where

$$\phi^x = D_x(\phi - \xi u_x) + \xi u_{xx} = 1 + u_x^2.$$

Also,

$$\phi^{xx} = D_x \phi^x - u_{xx} D_x \xi = 3u_x u_{xx},$$

thus the infinitesimal generator of the second prolongation $\mathbf{pr}^{(2)}SO(2)$ acting on $X \times U^{(2)}$ is

$$\mathbf{pr}^{(2)}v = -u \frac{\partial}{\partial x} + x \frac{\partial}{\partial u} + (1 + u_x^2) \frac{\partial}{\partial u_x} + 3u_x u_{xx} \frac{\partial}{\partial u_{xx}}.$$

Let v and w be two smooth vector-fields on $M \subset X \times U$. Then their n th prolongations, $\mathbf{pr}^{(n)}v$ and $\mathbf{pr}^{(n)}w$ respectively, have the *linearity property*

$$\mathbf{pr}^{(n)}(c_1v + c_2w) = c_1\mathbf{pr}^{(n)}v + c_2\mathbf{pr}^{(n)}w, \quad (c_1, c_2 - \text{constant}),$$

and the *Lie bracket property*

$$\mathbf{pr}^{(n)}[v, w] = [\mathbf{pr}^{(n)}v, \mathbf{pr}^{(n)}w].$$

3.9.3 Generalized Lie Symmetries

Consider a vector-field (3.70) defined on an open subset $M \subset X \times U$. Provided the coefficient functions ξ^i and ϕ^α depend only on x and u , v will generate a (local) one-parameter group of transformations $\exp(\varepsilon v)$ acting pointwise on the underlying space M . A significant generalization of the notion of symmetry group is get by relaxing this geometrical assumption, and allowing the coefficient functions ξ^i and ϕ^α to also depend on derivatives of u [Olver (1986)].

A *generalized vector-field* is a (formal) expression

$$v = \xi^i[u] \frac{\partial}{\partial x^i} + \phi^\alpha[u] \frac{\partial}{\partial u^\alpha}, \quad (i = 1, \dots, p, \alpha = 1, \dots, q), \quad (3.73)$$

in which ξ^i and ϕ^α are smooth functions. For example,

$$v = xu_x \frac{\partial}{\partial x} + u_{xx} \frac{\partial}{\partial u}$$

is a generalized vector in the case $p = q = 1$.

According to the general prolongation formula (3.71), we can define the *prolonged generalized vector-field*

$$\mathbf{pr}^{(n)}v = v + \phi_J^\alpha[u] \frac{\partial}{\partial u_J^\alpha},$$

whose coefficients are as before determined by the formula (3.72). Thus, in our previous example [Olver (1986)],

$$\mathbf{pr}^{(n)}v = xu_x \frac{\partial}{\partial x} + u_{xx} \frac{\partial}{\partial u} + [u_{xxx} - (xu_{xx} + u_x)u_x] \frac{\partial}{\partial u_x}.$$

Given a generalized vector-field v , its *infinite prolongation* (including all the derivatives) is the formal expression

$$\mathbf{pr} v = \xi^i \frac{\partial}{\partial x^i} + \phi_J^\alpha \frac{\partial}{\partial u_J^\alpha}.$$

Now, a generalized vector-field v is a *generalized infinitesimal symmetry* of a system \mathcal{S} of differential equations

$$\Delta_r[u] = \Delta_r(x, u^{(n)}) = 0, \quad (r = 1, \dots, l),$$

iff

$$\mathbf{pr} v[\Delta_r] = 0$$

for every smooth solution $u = f(x)$ [Olver (1986)].

For example, consider the heat equation

$$\Delta[u] = u_t - u_{xx} = 0.$$

The generalized vector-field $v = u_x \frac{\partial}{\partial u}$ has prolongation

$$\mathbf{pr} v = u_x \frac{\partial}{\partial u} + u_{xx} \frac{\partial}{\partial u_x} + u_{xt} \frac{\partial}{\partial u_t} + u_{xxx} \frac{\partial}{\partial u_{xx}} + \dots$$

Thus

$$\mathbf{pr} v(\Delta) = u_{xt} - u_{xxx} = D_x(u_t - u_{xx}) = D_x \Delta,$$

and hence v is a generalized symmetry of the heat equation.

3.9.3.1 Noether Symmetries

Here we present some results about *Noether symmetries*, in particular for the first-order Lagrangians $L(q, \dot{q})$ (see [Batlle *et. al.* (1989); Pons *et. al.* (2000)]). We start with a *Noether-Lagrangian symmetry*,

$$\delta L = \dot{F},$$

and we will investigate the conversion of this symmetry to the Hamiltonian formalism. Defining

$$G = (\partial L / \partial \dot{q}^i) \delta q^i - F,$$

we can write

$$\delta_i L \delta q^i + \dot{G} = 0, \tag{3.74}$$

where $\delta_i L$ is the *Euler-Lagrangian functional derivative* of L ,

$$\delta_i L = \alpha_i - W_{ik} \ddot{q}^k,$$

where

$$W_{ik} \equiv \frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^k} \quad \text{and} \quad \alpha_i \equiv -\frac{\partial^2 L}{\partial \dot{q}^i \partial q^k} \dot{q}^k + \frac{\partial L}{\partial \dot{q}^i}.$$

We consider the general case where the mass matrix, or *Hessian* (W_{ij}), may be a singular matrix. In this case there exists a kernel for the pull-back $\mathbb{F}L^*$ of the *Legendre map*, i.e., *fibre-derivative* $\mathbb{F}L$, from the *velocity phase-space manifold* TM (tangent bundle of the biodynamical manifold M) to the *momentum phase-space manifold* T^*M (cotangent bundle of M). This kernel is spanned by the vector-fields

$$\Gamma_\mu = \gamma_\mu^i \frac{\partial}{\partial \dot{q}^i},$$

where γ_μ^i are a basis for the null vectors of W_{ij} . The Lagrangian time-evolution differential operator can therefore be expressed as:

$$X = \partial_t + \dot{q}^k \frac{\partial}{\partial q^k} + a^k(q, \dot{q}) \frac{\partial}{\partial \dot{q}^k} + \lambda^\mu \Gamma_\mu \equiv X_o + \lambda^\mu \Gamma_\mu,$$

where a^k are functions which are determined by the formalism, and λ^μ are arbitrary functions. It is not necessary to use the Hamiltonian technique to find the Γ_μ , but it does facilitate the calculation:

$$\gamma_\mu^i = \mathbb{F}L^* \left(\frac{\partial \phi_\mu}{\partial p_i} \right), \quad (3.75)$$

where the ϕ_μ are the Hamiltonian primary first class constraints.

Notice that the highest derivative in (3.74), \ddot{q}^i , appears linearly. Because δL is a symmetry, (3.74) is identically satisfied, and therefore the coefficient of \ddot{q}^i vanishes:

$$W_{ik} \delta q^k - \frac{\partial G}{\partial \dot{q}^i} = 0. \quad (3.76)$$

We contract with a null vector γ_μ^i to find that

$$\Gamma_\mu G = 0.$$

It follows that G is projectable to a function G_H in T^*Q ; that is, it is the pull-back of a function (not necessarily unique) in T^*Q :

$$G = \mathbb{F}L^*(G_H).$$

This important property is valid for any conserved quantity associated with a Noether symmetry. Observe that G_H is determined up to the addition of linear combinations of the primary constraints. Substitution of this result in (3.76) gives

$$W_{ik} \left[\delta q^k - \mathbb{F}L^* \left(\frac{\partial G_H}{\partial p_k} \right) \right] = 0,$$

and so the brackets enclose a null vector of W_{ik} :

$$\delta q^i - \mathbb{F}L^* \left(\frac{\partial G_H}{\partial p_i} \right) = r^\mu \gamma_\mu^i, \quad (3.77)$$

for some $r^\mu(t, q, \dot{q})$.

We shall investigate the projectability of variations generated by diffeomorphisms in the following section. Assume that an infinitesimal transformation δq^i is projectable:

$$\Gamma_\mu \delta q^i = 0.$$

If δq^i is projectable, so must be r^μ , so that $r^\mu = \mathbb{F}L^*(r_H^\mu)$. Then, using (3.75) and (3.77), we see that

$$\delta q^i = \mathbb{F}L^* \left(\frac{\partial(G_H + r_H^\mu \phi_\mu)}{\partial p_i} \right).$$

We now redefine G_H to absorb the piece $r_H^\mu \phi_\mu$, and from now on we will have

$$\delta q^i = \mathbb{F}L^* \left(\frac{\partial G_H}{\partial p_i} \right).$$

Define

$$\hat{p}_i = \frac{\partial L}{\partial \dot{q}^i};$$

after eliminating (3.76) times \ddot{q}^i from (3.74), we get

$$\left(\frac{\partial L}{\partial q^i} - \dot{q}^k \frac{\partial \hat{p}_i}{\partial q^k} \right) \mathbb{F}L^* \left(\frac{\partial G_H}{\partial p_i} \right) + \dot{q}^i \frac{\partial}{\partial q^i} \mathbb{F}L^*(G_H) + \mathbb{F}L^* \partial_t G_H = 0,$$

which simplifies to

$$\frac{\partial L}{\partial q^i} \mathbb{F}L^* \left(\frac{\partial G_H}{\partial p_i} \right) + \dot{q}^i \mathbb{F}L^* \left(\frac{\partial G_H}{\partial q^i} \right) + \mathbb{F}L^* \partial_t G_H = 0. \quad (3.78)$$

Now let us invoke two identities [Batlle *et. al.* (1989)] that are at the core of the connection between the Lagrangian and the Hamiltonian equations of motion. They are

$$\dot{q}^i = \mathbb{F}L^*\left(\frac{\partial H}{\partial p_i}\right) + v^\mu(q, \dot{q})\mathbb{F}L^*\left(\frac{\partial \phi_\mu}{\partial p_i}\right),$$

and

$$\frac{\partial L}{\partial \dot{q}^i} = -\mathbb{F}L^*\left(\frac{\partial H}{\partial q^i}\right) - v^\mu(q, \dot{q})\mathbb{F}L^*\left(\frac{\partial \phi_\mu}{\partial q^i}\right);$$

where H is any canonical Hamiltonian, so that $\mathbb{F}L^*(H) = \dot{q}^i(\partial L/\partial \dot{q}^i) - L = \hat{E}$, the Lagrangian energy, and the functions v^μ are determined so as to render the first relation an identity. Notice the important relation

$$\Gamma_\mu v^\nu = \delta_\mu^\nu,$$

which stems from applying Γ_μ to the first identity and taking into account that

$$\Gamma_\mu \circ \mathbb{F}L^* = 0.$$

Substitution of these two identities into (3.78) induces (where $\{ , \}$ denotes the *Poisson bracket*)

$$\mathbb{F}L^*\{G_H, H\} + v^\mu \mathbb{F}L^*\{G_H, \phi_\mu\} + \mathbb{F}L^*\partial_t G_H = 0.$$

This result can be split through the action of Γ_μ into

$$\mathbb{F}L^*\{G_H, H\} + \mathbb{F}L^*\partial_t G_H = 0,$$

and

$$\mathbb{F}L^*\{G_H, \phi_\mu\} = 0;$$

or equivalently,

$$\{G_H, H\} + \partial_t G_H = pc,$$

and

$$\{G_H, \phi_\mu\} = pc,$$

where pc stands for any linear combination of primary constraints. In this way, we have arrived at a neat characterization for a generator G_H of Noether transformations in the canonical formalism.

3.9.4 Application: *Biophysical PDEs*

In this subsection we consider two most important equations for biophysics:

- (1) The heat equation, which has been analyzed in muscular mechanics since the early works of A.V. Hill ([Hill (1938)]); and
- (2) The Korteveg–de Vries equation, the basic equation for solitary models of muscular excitation–contraction dynamics.

Suppose

$$\mathcal{S} : \Delta_r(x, u^{(n)}) = 0, \quad (r = 1, \dots, l),$$

is a system of DEs of maximal rank defined over $M \subset X \times U$. If G is a local group of transformations acting on M , and

$$\mathbf{pr}^{(n)}v[\Delta_r(x, u^{(n)})] = 0, \quad \text{whenever} \quad \Delta(x, u^{(n)}) = 0, \quad (3.79)$$

(with $r = 1, \dots, l$) for every infinitesimal generator v of G , then G is a symmetry group of the system \mathcal{S} [Olver (1986)].

3.9.4.1 The Heat Equation

Recall that the $(1 + 1)$ D *heat equation* (with the thermal diffusivity normalized to unity)

$$u_t = u_{xx} \quad (3.80)$$

has two independent variables x and t , and one dependent variable u , so $p = 2$ and $q = 1$. Equation (3.80) has the second-order, $n = 2$, and can be identified with the linear submanifold $M^{(2)} \subset X \times U^{(2)}$ determined by the vanishing of $\Delta(x, t, u^{(2)}) = u_t - u_{xx}$.

Let

$$v = \xi(x, t, u) \frac{\partial}{\partial x} + \tau(x, t, u) \frac{\partial}{\partial t} + \phi(x, t, u) \frac{\partial}{\partial u}$$

be a vector-field on $X \times U$. According to (3.79) we need to now the second prolongation

$$\mathbf{pr}^{(2)}v = v + \phi^x \frac{\partial}{\partial u_x} + \phi^t \frac{\partial}{\partial u_t} + \phi^{xx} \frac{\partial}{\partial u_{xx}} + \phi^{xt} \frac{\partial}{\partial u_{xt}} + \phi^{tt} \frac{\partial}{\partial u_{tt}}$$

of v . Applying $\mathbf{pr}^{(2)}v$ to (3.80) we find the infinitesimal criterion (3.79) to be

$$\phi^t = \phi^{xx},$$

which must be satisfied whenever $u_t = u_{xx}$.

3.9.4.2 The Kortevég-De Vries Equation

Recall that the *Kortevég-de Vries equation*

$$u_t + u_{xxx} + uu_x = 0 \quad (3.81)$$

arises in physical systems in which both nonlinear and dispersive effects are relevant. A vector-field

$$v = \xi(x, t, u) \frac{\partial}{\partial x} + \tau(x, t, u) \frac{\partial}{\partial t} + \phi(x, t, u) \frac{\partial}{\partial u}$$

generates a one-parameter symmetry group iff

$$\phi^t + \phi^{xxx} + u\phi^x + u_x\phi = 0,$$

whenever u satisfies (3.81), etc.

3.9.5 Lie-Invariant Geometric Objects

3.9.5.1 Robot Kinematics

It is well known (see [Blackmore and Leu (1992); Prykarpatsky (1996)]), that motion planning, numerically controlled machining and robotics are just a few of many areas of manufacturing automation in which the analysis and representation of *swept volumes* plays a crucial role. The swept volume modelling is also an important part of task-oriented robot motion planning. A typical *motion planning* problem consists in a collection of objects moving around obstacles from an initial to a final configuration. This may include in particular, solving the *collision detection* problem.

When a solid object undergoes a rigid motion, the totality of points through which it passed constitutes a region in space called the *swept volume*. To describe the geometrical structure of the swept volume we pose this problem as one of geometric study of some manifold swept by surface points using powerful tools from both modern differential geometry and nonlinear dynamical systems theory [Ricca (1993); Langer and Perline (1994); Prykarpatsky (1996); Groesen and Jager (1994)] on manifolds. For some special cases of the Euclidean motion in the space \mathbb{R}^3 one can construct a very rich hydrodynamic system [Blackmore and Leu (1992)] modelling a sweep flow, which appears to be a *completely integrable Hamiltonian system* having a special *Lax type representation*. To describe in detail these and

other properties of swept volume dynamical systems, we develop *Cartan's theory of Lie-invariant geometric objects* generated by closed ideals in the Grassmann's algebra, following [Blackmore *et. al.* (1998)].

Let a Lie group G act on an analytical manifold Y in the transitive way, that is the action $G \times Y \xrightarrow{\rho} Y$ generates some nonlinear exact representation of the Lie group G on the manifold Y . In the frame of the Cartan's theory, the representation $G \times Y \xrightarrow{\rho} Y$ can be described by means of a system of differential 1-forms (see section 5.8 below)

$$\bar{\beta}^j = dy^j + \xi_i^j \bar{\omega}^i(a, da) \quad (3.82)$$

in the Grassmann algebra $\Lambda(Y \times G)$ on the product $Y \times G$, where $\bar{\omega}^i(a, da) \in T_a^*(G)$, $i = 1, \dots, r = \dim G$ is a basis of left-invariant Cartan's forms of the Lie group G at a point $a \in G$, $y = \{y^j : j = 1, \dots, n = \dim Y\} \in Y$ and $\xi_i^j : Y \times G \rightarrow \mathbb{R}$ are some smooth real valued functions.

The following *Cartan Theorem* (see [Blackmore *et. al.* (1998)]) is basic in describing a geometric object invariant with respect to the mentioned above group action $G \times Y \xrightarrow{\rho} Y$: The system of differential forms (3.82) is a system of an invariant geometric object iff the following conditions are fulfilled:

- (1) The coefficients $\xi_i^j \in C^k(Y; \mathbb{R})$ for all $i = 1, \dots, r$, $j = 1, \dots, n$, are some *analytical functions* on Y ; and
- (2) The differential system (3.82) is completely integrable within the *Frobenius–Cartan criterion*.

The Cartan's Theorem actually says that the differential system (3.82) can be written down as

$$\bar{\beta}^j = dy^j + \xi_i^j(y) \bar{\omega}^i(a, da), \quad (3.83)$$

where 1-forms $\{\bar{\omega}^i(a, da) : i = 1, \dots, r\}$ satisfy the standard *Maurer–Cartan equations*

$$\bar{\Omega}^j = d\bar{\omega}^j + \frac{1}{2} c_{ik}^j \bar{\omega}^i \wedge \bar{\omega}^k = 0, \quad (3.84)$$

for all $j = 1, \dots, r$ on G , coefficients $c_{ik}^j \in \mathbb{R}$, $i, j, k = 1, \dots, r$, being the corresponding structure constants of the Lie algebra \mathcal{G} of the Lie group G .

3.9.5.2 Maurer–Cartan 1-Forms

Let be given a Lie group G with the Lie algebra $\mathcal{G} \simeq T_e(G)$, whose basis is a set $\{A_i \in \mathcal{G} : i = 1, \dots, r\}$, where $r = \dim G \equiv \dim \mathcal{G}$. Let also a set $U_0 \subset \{a^i \in \mathbb{R} : i = 1, \dots, r\}$ be some open neighborhood of the zero point in \mathbb{R}^r . The exponential mapping $\exp : U_0 \rightarrow G_0$, where by definition [Blackmore *et. al.* (1998)]

$$\mathbb{R}^r \supset U_0 \ni (a^1, \dots, a^r) : \xrightarrow{\exp} \exp(a^i A_i) = a \in G_0 \subset G, \quad (3.85)$$

is an analytical mapping of the whole U_0 on some open neighborhood G_0 of the unity element $e \in G$. From (3.85) it is easy to find that $T_e(G) = T_e(G_0) \simeq \mathcal{G}$, where $e = \exp(0) \in G$. Define now the following left-invariant \mathcal{G} -valued differential 1-form on $G_0 \subset G$:

$$\bar{\omega}(a, da) = a^{-1} da = \bar{\omega}^j(a, da) A_j, \quad (3.86)$$

where $A_j \in \mathcal{G}$, $\bar{\omega}^j(a, da) \in T_a^*(G)$, $a \in G_0$, $j = 1, \dots, r$. To build effectively the unknown forms $\{\bar{\omega}^j(a, da) : j = 1, \dots, r\}$, let us consider the following analytical one-parameter 1-form $\bar{\omega}_t(a, da) = \bar{\omega}(a_t; da_t)$ on G_0 , where $a_t = \exp(ta^i A_i)$, $t \in [0, 1]$, and differentiate this form with respect to the parameter $t \in [0, 1]$. We will get [Blackmore *et. al.* (1998)]

$$d\bar{\omega}_t/dt = -a^j A_j a_t^{-1} da_t + a_t^{-1} a_t da_t^j A_j + a_t^{-1} da_t a^j A_j = -a^j [A_j, \bar{\omega}_t] + A_j da_j. \quad (3.87)$$

Having used the Lie identity $[A_j, A_k] = c_{jk}^i A_i$, $j, k = 1, \dots, r$, and the right hand side of (3.86) in form

$$\bar{\omega}^j(a, da) = \bar{\omega}_k^j(a) da^k, \quad (3.88)$$

we ultimately get that

$$\frac{d}{dt}(t\bar{\omega}_i^j(ta)) = \mathcal{A}_k^j t\bar{\omega}_i^k(ta) + \delta_i^j, \quad (3.89)$$

where the matrix \mathcal{A}_i^k , $i, k = 1, \dots, r$, is defined as follows:

$$\mathcal{A}_i^k = c_{ij}^k a^j. \quad (3.90)$$

Thus, the matrix $W_i^j(t) = t\bar{\omega}_i^j(ta)$, $i, j = 1, \dots, r$, satisfies the following from (3.89) differential equation [Chevalley (1955)]

$$dW/dt = \mathcal{A}W + E, \quad W|_{t=0} = 0, \quad (3.91)$$

where $E = \|\delta_i^j\|$ is the unity matrix. The solution of (3.91) is representable as

$$W(t) = \sum_{n=1}^{\infty} \frac{t^n}{n!} \mathcal{A}^{n-1} \quad (3.92)$$

for all $t \in [0, 1]$. Whence, recalling the above definition of the matrix $W(t)$, we get easily that

$$\bar{\omega}_k^j(a) = W_k^j(t) \Big|_{t=1} = \sum_{n=1}^{\infty} (n!)^{-1} \mathcal{A}^{n-1}. \quad (3.93)$$

Therefore, the following Theorem solves the problem of finding in an effective algebraic way corresponding to a Lie algebra \mathcal{G} the left-invariant 1-form $\bar{\omega}(a, da) \in T_a^*(G) \otimes \mathcal{G}$ at any $a \in G$: Let's be given a Lie algebra G with the structure constants $c_{ij}^k \in R$, $i, j, k = 1, \dots, r = \dim \mathcal{G}$, related to some basis $\{A_j \in G : j = 1, \dots, r\}$. Then the adjoint to G left-invariant Maurer–Cartan 1-form $\bar{\omega}(a, da)$ is built as follows [Blackmore *et. al.* (1998)]:

$$\bar{\omega}(a, da) = A_j \bar{\omega}_k^j(a) da^k, \quad (3.94)$$

where the matrix $W = \|\bar{\omega}_k^j(a)\|$, $j, k = 1, \dots, r$, is given exactly as

$$W = \sum_{n=1}^{\infty} (n!)^{-1} \mathcal{A}^{n-1}, \quad \mathcal{A}_k^j = c_{ki}^j a^i. \quad (3.95)$$

Below we shall try to use the experience gained above in solving an analogous problem of the theory of connections over a principal fibre bundle $P(M; G)$ as well as over associated with it a fibre bundle $P(M; Y, G)$.

3.9.5.3 General Structure of Integrable One-Forms

Given 2-forms generating a closed ideal $\mathfrak{I}(\alpha)$ in the Grassmann algebra $\Lambda(M)$, we will denote as above by $\mathfrak{I}(\alpha, \beta)$ an augmented ideal in $\Lambda(M; Y)$, where the manifold Y will be called in further the representation space of some adjoint Lie group G action: $G \times Y \xrightarrow{\rho} Y$. We can find, therefore, the determining relationships for the set of 1-forms $\{\beta\}$ and 2-forms $\{\alpha\}$

$$\begin{aligned} \{\alpha\} &= \{\alpha^j \in \Lambda^2(M) : j = 1, \dots, m_\alpha\}, \\ \{\beta\} &= \{\beta^j \in \Lambda^1(M \times Y) : j = 1, \dots, n = \dim Y\}, \end{aligned} \quad (3.96)$$

satisfying such equations [Blackmore *et. al.* (1998)]:

$$\begin{aligned} d\alpha^i &= a_k^i(\alpha) \wedge \alpha^k, \\ d\beta^j &= f_k^j \alpha^k + \omega_s^j \wedge \beta^s, \end{aligned} \quad (3.97)$$

where $a_k^i(\alpha) \in \Lambda^1(M)$, $f_k^j \in \Lambda^0(M \times Y)$ and $\omega_s^j \in \Lambda^1(M \times Y)$ for all $i, k = 1, \dots, m_\alpha$, $j, s = 1, \dots, n$. Since the identity $d^2\beta^j \equiv 0$ takes place for all $j = 1, \dots, n$, from (3.97) we deduce the following relationship:

$$\left(d\omega_k^j + \omega_s^j \wedge \omega_k^s \right) \wedge \beta^k + \left(df_s^j + \omega_k^j f_s^k + f_l^j a_s^l(\alpha) \right) \wedge \alpha^s \equiv 0. \quad (3.98)$$

As a result of (3.98) we get [Blackmore *et. al.* (1998)]

$$\begin{aligned} d\omega_k^j + \omega_s^j \wedge \omega_k^s &\in \mathfrak{I}(\alpha, \beta), \\ df_s^j + \omega_k^j f_s^k + f_l^j a_s^l(\alpha) &\in \mathfrak{I}(\alpha, \beta) \end{aligned} \quad (3.99)$$

for all $j, k = 1, \dots, n$, $s = 1, \dots, m_\alpha$. The second inclusion in (3.99) gives a possibility to define the 1-forms $\theta_s^j = f_l^j a_s^l(\alpha)$ satisfying the inclusion

$$d\theta_s^j + \omega_k^j \wedge \theta_s^k \in \mathfrak{I}(\alpha, \beta) \oplus f_l^j c_s^l(\alpha), \quad (3.100)$$

which we obtained having used the identities $d^2\alpha^j \equiv 0$, $j = 1, \dots, m_\alpha$, in the form $c_s^j(\alpha) \wedge \alpha^s \equiv 0$,

$$c_s^j(\alpha) = da_s^j(\alpha) + a_l^j(\alpha) \wedge a_s^l(\alpha), \quad (3.101)$$

following from (3.97). Let us suppose further that as $s = s_0$ the 2-forms $c_{s_0}^j(\alpha) \equiv 0$ for all $j = 1, \dots, m_\alpha$. Then as $s = s_0$, we can define a set of 1-forms $\theta^j = \theta_{s_0}^j \in \Lambda^1(M \times Y)$, $j = 1, \dots, n$, satisfying the exact inclusions:

$$d\theta^j + \omega_k^j \wedge \theta^k = \Theta^j \in \mathfrak{I}(\alpha, \beta) \quad (3.102)$$

together with a set of inclusions for 1-forms $\omega_k^j \in \Lambda^1(M \times Y)$

$$d\omega_k^j + \omega_s^j \wedge \omega_k^s = \Omega_k^j \in \mathfrak{I}(\alpha, \beta) \quad (3.103)$$

As it follows from the general theory [Sulanke and Wintgen (1972)] of connections on the fibred frame space $P(M; GL(n))$ over a base manifold M , we can interpret the equations (3.103) as the equations defining the curvature 2-forms $\Omega_k^j \in \Lambda^2(P)$, as well as interpret the equations (3.102) as those, defining the torsion 2-forms $\Theta^j \in \Lambda^2(P)$. Since $\mathfrak{I}(\alpha) = 0 = \mathfrak{I}(\alpha, \beta)$ upon the integral submanifold $\bar{M} \subset M$, the reduced fibred frame space $P(\bar{M}; GL(n))$ will have the flat curvature and be torsion free, being as a

result, completely trivialized on $\bar{M} \subset M$. Consequently, we can formulate the following Theorem.

Let the condition above on the ideals $\mathfrak{I}(\alpha)$ and $\mathfrak{I}(\alpha, \beta)$ be fulfilled. Then the set of 1-forms $\{\beta\}$ generates the integrable augmented ideal $\mathfrak{I}(\alpha, \beta) \subset \Lambda(M \times Y)$ iff there exists some curvature 1-form $\omega \in \Lambda^1(P) \otimes \mathfrak{gl}(n)$ and torsion 1-form $\theta \in \Lambda^1(P) \otimes \mathbb{R}^n$ on the adjoint fibred frame space $P(M; GL(n))$, satisfying the inclusions [Blackmore *et. al.* (1998)]

$$\begin{aligned} d\omega + \omega \wedge \omega &\in \mathfrak{I}(\alpha, \beta) \otimes \mathfrak{gl}(n), \\ d\theta + \omega \wedge \theta &\in \mathfrak{I}(\alpha, \beta) \otimes \mathbb{R}^n. \end{aligned} \quad (3.104)$$

Upon the reduced fibred frame space $P(\bar{M}; GL(n))$ the corresponding curvature and torsion are vanishing, where $\bar{M} \subset M$ is the integral submanifold of the ideal $\mathfrak{I}(\alpha) \subset \Lambda(M)$.

3.9.5.4 Lax Integrable Dynamical Systems

Consider some set $\{\beta\}$ defining a Cartan's Lie group G invariant object on a manifold $M \times Y$:

$$\beta^j = dy^j + \xi_k^j(y)b^k(z), \quad (3.105)$$

where $i = 1, \dots, n = \dim Y$, $r = \dim G$. The set (3.105) defines on the manifold Y a set $\{\xi\}$ of vector-fields, compiling a representation $\rho: \mathcal{G} \rightarrow \{\xi\}$ of a given Lie algebra \mathcal{G} , that is vector-fields $\xi_s = \xi_s^j(y) \frac{\partial}{\partial y^j} \in \{\xi\}$, $s = 1, \dots, r$, enjoy the following Lie algebra \mathcal{G} relationships

$$[\xi_s, \xi_l] = c_{sl}^k \xi_k \quad (3.106)$$

for all $s, l, k = 1, \dots, r$. We can now compute the differentials $d\beta^j \in \Lambda^2(M \times Y)$, $j = 1, \dots, n$, using (3.105) and (3.106) as follows [Blackmore *et. al.* (1998)]:

$$d\beta^j = \frac{\partial \xi_k^j(y)}{\partial y^l} \left(\beta^l - \xi_s^l(y)b^s(z) \right) \wedge b^k(z) + \xi_k^j(y) db^k(z) \quad (3.107)$$

which is equal to

$$\frac{\partial \xi_k^j(y)}{\partial y^l} \beta^l \wedge b_k(z) + \xi_l^j \left(db^l(z) + \frac{1}{2} c_{ks}^l db^k(z) \wedge db^s(z) \right),$$

where $\{\alpha\} \subset \Lambda^2(M)$ is some *a priori* given integrable system of 2-forms on M , vanishing upon the integral submanifold $\bar{M} \subset M$. It is obvious that

inclusions (3.107) take place iff the following conditions are fulfilled: for all $j = 1, \dots, r$

$$db^j(z) + \frac{1}{2}c_{ks}^j db^k(z) \wedge db^s(z) \in \mathfrak{I}(\alpha). \quad (3.108)$$

The inclusions (3.108) mean in particular, that upon the integral submanifold $\bar{M} \subset M$ of the ideal $\mathfrak{I}(\alpha) \subset \Lambda(M)$ the equalities

$$\mu^* \bar{\omega}^j \equiv s^* b^j \quad (3.109)$$

are true, where $\bar{\omega}^j \in T_e^*(G)$, $j = 1, \dots, r$, are the left-invariant Maurer–Cartan forms on the invariance Lie group G . Thus, due to inclusions (3.108) all conditions of Cartan’s Theorem are enjoyed, giving rise to a possibility to get the set of forms $b^j(z) \in \Lambda^1(M)$ in an explicit form. To do this, let us define a \mathcal{G} -valued curvature 1-form $\omega \in \Lambda^1(P(M; G)) \otimes \mathcal{G}$ as follows [Blackmore *et. al.* (1998)]

$$\omega = Ad_{a^{-1}}(A_j b^j) + \bar{\omega} \quad (3.110)$$

where $\bar{\omega} \in \mathcal{G}$ is the standard Maurer–Cartan 1-form on G . This 1-form satisfies followed by (3.108) the canonical structure inclusion for $\Gamma = A_j b^j \in \Lambda^1(M) \otimes \mathcal{G}$:

$$d\Gamma + \Gamma \wedge \Gamma \in \mathfrak{I}(\alpha) \otimes \mathcal{G}, \quad (3.111)$$

serving as a main relationships determining the form (3.110).

3.9.5.5 APPLICATION: Burgers Dynamical System

Consider the *Burgers dynamical system* on a *functional manifold* $M \subset C^k(\mathbb{R}; \mathbb{R})$:

$$u_t = uu_x + u_{xx}, \quad (3.112)$$

where $u \in M$ and $t \in \mathbb{R}$ is an evolution (time) parameter. The flow of (3.112) on M can be recast into a set of 2-forms $\{\alpha\} \subset \Lambda^2(J(\mathbb{R}^2; \mathbb{R}))$ upon the adjoint jet-manifold $J(\mathbb{R}^2; \mathbb{R})$ (see section 5.8 below) as follows [Blackmore *et. al.* (1998)]:

$$\begin{aligned} \{\alpha\} = \{ & du^{(0)} \wedge dt - u^{(1)} dx \wedge dt = \alpha^1, \quad du^{(0)} \wedge dx + u^{(0)} du^{(0)} \wedge dt \\ & + du^{(1)} \wedge dt = \alpha^2 : (x, t; u^{(0)}, u^{(1)})^\tau \in M^4 \subset J^1(\mathbb{R}^2; \mathbb{R}) \}, \end{aligned} \quad (3.113)$$

where M^4 is a 4-D submanifold in $J^1(\mathbb{R}^2; \mathbb{R})$ with coordinates $(x, t, u^{(0)} = u, u^{(1)} = u_x)$. The set of 2-forms (3.113) generates the closed ideal $\mathfrak{I}(\alpha)$, since

$$d\alpha^1 = dx \wedge \alpha^2 - u^{(0)} dx \wedge \alpha^1, \quad d\alpha^2 = 0, \quad (3.114)$$

the integral submanifold $\bar{M} = \{x, t \in \mathbb{R}\} \subset M^4$ being defined by the condition $\mathfrak{I}(\alpha) = 0$. We now look for a reduced ‘curvature’ 1-form $\Gamma \in \Lambda^1(M^4) \otimes \mathfrak{g}$, belonging to some (not yet determined) Lie algebra \mathfrak{g} . This 1-form can be represented using (3.113), as follows:

$$\Gamma = b^{(x)}(u^{(0)}, u^{(1)})dx + b^{(t)}(u^{(0)}, u^{(1)})dt, \quad (3.115)$$

where elements $b^{(x)}, b^{(t)} \in \mathfrak{g}$ satisfy [Blackmore *et. al.* (1998)]

$$\begin{aligned} \frac{\partial b^{(x)}}{\partial u^{(0)}} &= g_2, & \frac{\partial b^{(x)}}{\partial u^{(1)}} &= 0, & \frac{\partial b^{(t)}}{\partial u^{(0)}} &= g_1 + g_2 u^{(0)}, \\ \frac{\partial b^{(t)}}{\partial u^{(1)}} &= g_2, & [b^{(x)}, b^{(t)}] &= -u^{(1)} g_1. \end{aligned} \quad (3.116)$$

The set (3.116) has the following unique solution

$$b^{(x)} = A_0 + A_1 u^{(0)}, \quad b^{(t)} = u^{(1)} A_1 + \frac{u^{(0)2}}{2} A_1 + [A_1, A_0] u^{(0)} + A_2, \quad (3.117)$$

where $A_j \in \mathfrak{g}$, $j = \overline{0, 2}$, are some constant elements on M of a Lie algebra \mathfrak{g} under search, satisfying the next Lie structure equations:

$$\begin{aligned} [A_0, A_2] &= 0, \\ [A_0, [A_1, A_0]] + [A_1, A_2] &= 0, \\ [A_1, [A_1, A_0]] + \frac{1}{2}[A_0, A_1] &= 0. \end{aligned} \quad (3.118)$$

From (3.116) one can see that the curvature 2-form $\Omega \in \text{span}_{\mathbb{R}}\{A_1, [A_0, A_1] : A_j \in \mathfrak{g}, j = \overline{0, 1}\}$. Therefore, reducing via the *Ambrose–Singer Theorem* the associated principal fibered frame space $P(M; G = GL(n))$ to the principal fibre bundle $P(M; G(h))$, where $G(h) \subset G$ is the corresponding holonomy Lie group of the connection Γ on P , we need to satisfy the following conditions for the set $\mathfrak{g}(h) \subset \mathfrak{g}$ to be a Lie subalgebra in \mathfrak{g} : $\nabla_x^m \nabla_t^n \Omega \in \mathfrak{g}(h)$ for all $m, n \in \mathbb{Z}_+$.

Let us try now to close the above procedure requiring that [Blackmore *et. al.* (1998)]

$$\mathfrak{g}(h) = \mathfrak{g}(h)_0 = \text{span}_{\mathbb{R}}\{\nabla_x^m \nabla_t^n \Omega \in \mathfrak{g} : m + n = 0\} \quad (3.119)$$

This means that

$$\mathfrak{g}(h)_0 = \text{span}_{\mathbb{R}}\{A_1, A_3 = [A_0, A_1]\}. \quad (3.120)$$

To satisfy the set of relations (3.118) we need to use expansions over the basis (3.120) of the external elements $A_0, A_2 \in \mathfrak{g}(h)$:

$$A_0 = q_{01}A_1 + q_{13}A_3, \quad A_2 = q_{21}A_1 + q_{23}A_3. \quad (3.121)$$

Substituting expansions (3.121) into (3.118), we get that $q_{01} = q_{23} = \lambda$, $q_{21} = -\lambda^2/2$ and $q_{03} = -2$ for some arbitrary real parameter $\lambda \in \mathbb{R}$, that is $\mathfrak{g}(h) = \text{span}_{\mathbb{R}}\{A_1, A_3\}$, where

$$[A_1, A_3] = A_3/2; \quad A_0 = \lambda A_1 - 2A_3, \quad A_2 = -\lambda^2 A_1/2 + \lambda A_3. \quad (3.122)$$

As a result of (3.122) we can state that the holonomy Lie algebra $\mathfrak{g}(h)$ is a real 2D one, assuming the following (2×2) -matrix representation [Blackmore *et. al.* (1998)]:

$$\begin{aligned} A_1 &= \begin{pmatrix} 1/4 & 0 \\ 0 & -1/4 \end{pmatrix}, & A_3 &= \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \\ A_0 &= \begin{pmatrix} \lambda/4 & -2 \\ 0 & -\lambda/4 \end{pmatrix}, & A_2 &= \begin{pmatrix} -\lambda^2/8 & \lambda \\ 0 & \lambda^2/8 \end{pmatrix}. \end{aligned} \quad (3.123)$$

Thereby from (3.115), (3.117) and (3.123) we get the *reduced curvature 1-form* $\Gamma \in \Lambda^1(M) \otimes \mathfrak{g}$,

$$\Gamma = (A_0 + uA_1)dx + ((u_x + u^2/2)A_1 - uA_3 + A_2)dt, \quad (3.124)$$

generating *parallel transport* of vectors from the representation space Y of the holonomy Lie algebra $\mathfrak{g}(h)$:

$$dy + \Gamma y = 0, \quad (3.125)$$

upon the integral submanifold $\bar{M} \subset M^4$ of the ideal $\mathcal{I}(\alpha)$, generated by the set of 2-forms (3.113). The result (3.125) means also that the Burgers dynamical system (3.112) is endowed with the standard Lax type representation, having the spectral parameter $\lambda \in \mathbb{R}$ necessary for its integrability in quadratures.

3.10 Riemannian Manifolds and Their Applications

3.10.1 Local Riemannian Geometry

An important class of problems in Riemannian geometry is to understand the interaction between the curvature and topology on a smooth manifold (see [Cao and Chow (1999)]). A prime example of this interaction is the *Gauss–Bonnet formula* on a closed surface M^2 , which says

$$\int_M K \, dA = 2\pi \chi(M), \quad (3.126)$$

where dA is the area element of a metric g on M , K is the Gaussian curvature of g , and $\chi(M)$ is the *Euler characteristic* of M .

To study the geometry of a smooth manifold we need an additional structure: the *Riemannian metric tensor*. The metric is an inner product on each of the tangent spaces and tells us how to measure angles and distances infinitesimally. In local coordinates (x^1, x^2, \dots, x^n) , the metric g is given by $g_{ij}(x) \, dx^i \otimes dx^j$, where $(g_{ij}(x))$ is a positive definite symmetric matrix at each point x . For a smooth manifold one can differentiate functions. A Riemannian metric defines a natural way of differentiating vector-fields: *covariant differentiation*. In Euclidean space, one can change the order of differentiation. On a Riemannian manifold the commutator of twice covariant differentiating vector-fields is in general nonzero and is called the *Riemann curvature tensor*, which is a 4-tensor-field on the manifold.

For surfaces, the Riemann curvature tensor is equivalent to the *Gaussian curvature* K , a scalar function. In dimensions 3 or more, the Riemann curvature tensor is inherently a tensor-field. In local coordinates, it is denoted by R_{ijkl} , which is anti-symmetric in i and k and in j and l , and symmetric in the pairs $\{ij\}$ and $\{kl\}$. Thus, it can be considered as a bilinear form on 2-forms which is called the *curvature operator*. We now describe heuristically the various curvatures associated to the Riemann curvature tensor. Given a point $x \in M^n$ and 2-plane Π in the tangent space of M at x , we can define a surface S in M to be the union of all geodesics passing through x and tangent to Π . In a neighborhood of x , S is a smooth 2D submanifold of M . We define the *sectional curvature* $K(\Pi)$ of the 2-plane to be the Gauss curvature of S at x :

$$K(\Pi) = K_S(x).$$

Thus the sectional curvature K of a Riemannian manifold associates to each

2-plane in a tangent space a real number. Given a line L in a tangent space, we can average the sectional curvatures of all planes through L to get the *Ricci tensor* $Rc(L)$. Likewise, given a point $x \in M$, we can average the Ricci curvatures of all lines in the tangent space of x to get the *scalar curvature* $R(x)$. In local coordinates, the Ricci tensor is given by $R_{ik} = g^{jl} R_{ijkl}$ and the scalar curvature is given by $R = g^{ik} R_{ik}$, where $(g^{ij}) = (g_{ij})^{-1}$ is the inverse of the metric tensor (g_{ij}) .

3.10.1.1 Riemannian Metric on M

In this section we mainly follow [Petersen (1999); Petersen (1998)].

Riemann in 1854 observed that around each point $m \in M$ one can pick a *special* coordinate system (x^1, \dots, x^n) such that there is a symmetric $(0, 2)$ -tensor-field $g_{ij}(m)$ called the *metric tensor* defined as

$$g_{ij}(m) = g(\partial_{x^i}, \partial_{x^j}) = \delta_{ij}, \quad \partial_{x^k} g_{ij}(m) = 0.$$

Thus the metric, at the specified point $m \in M$, in the coordinates (x^1, \dots, x^n) looks like the Euclidean metric on \mathbb{R}^n . We emphasize that these conditions only hold at the specified point $m \in M$. When passing to different points it is necessary to pick different coordinates. If a curve γ passes through m , say, $\gamma(0) = m$, then the acceleration at 0 is defined by firstly, writing the curve out in our special coordinates

$$\gamma(t) = (\gamma^1(t), \dots, \gamma^n(t)),$$

secondly, defining the tangent, *velocity* vector-field, as

$$\dot{\gamma} = \dot{\gamma}^i(t) \cdot \partial_{x^i},$$

and finally, the *acceleration* vector-field as

$$\ddot{\gamma}(0) = \ddot{\gamma}^i(0) \cdot \partial_{x^i}.$$

Here, the background idea is that we have a *connection*.

Recall that a *connection* on a smooth manifold M tells us how to *parallel transport a vector* at a point $x \in M$ to a vector at a point $x' \in M$ along a curve $\gamma \in M$. Roughly, to parallel transport vectors along curves, it is enough if we can define parallel transport under an infinitesimal displacement: given a vector X at x , we would like to define its parallel transported version \tilde{X} after an infinitesimal displacement by ϵv , where v is a tangent vector to M at x .

More precisely, a vector-field X along a parameterized curve $\alpha : I \rightarrow M$ in M is *tangent to M along α* if $X(t) \in M_{\alpha(t)}$ for all for $t \in I \subset \mathbb{R}$. However, the derivative \dot{X} of such a vector-field is, in general, not tangent to M . We can, nevertheless, get a vector-field tangent to M by projecting $\dot{X}(t)$ orthogonally onto $M_{\alpha(t)}$ for each $t \in I$. This process of differentiating and then projecting onto the tangent space to M defines an operation with the same properties as differentiation, except that now differentiation of vector-fields tangent to M induces vector-fields tangent to M . This operation is called *covariant differentiation*.

Let $\gamma : I \rightarrow M$ be a parameterized curve in M , and let X be a smooth vector-field tangent to M along α . The *absolute covariant derivative* of X is the vector-field $\dot{\tilde{X}}$ tangent to M along α , defined by $\dot{\tilde{X}} = \dot{X}(t) - [\dot{X}(t) \cdot N(\alpha(t))] N(\alpha(t))$, where N is an orientation on M . Note that $\dot{\tilde{X}}$ is independent of the choice of N since replacing N by $-N$ has no effect on the above formula.

Lie bracket (3.7.2) defines a *symmetric affine connection* ∇ on any manifold M :

$$[X, Y] = \nabla_X Y - \nabla_Y X.$$

In case of a Riemannian manifold M , the connection ∇ is also compatible with the Riemannian metrics g on M and is called the *Levi-Civita connection* on TM .

For a function $f \in C^k(M, \mathbb{R})$ and a vector a vector-field $X \in \mathcal{X}^k(M)$ we always have the Lie derivative (3.7)

$$\mathcal{L}_X f = \nabla_X f = df(X).$$

But there is no natural definition for $\nabla_X Y$, where $Y \in \mathcal{X}^k(M)$, unless one also has a Riemannian metric. Given the tangent field $\dot{\gamma}$, the acceleration can then be computed by using a Leibniz rule on the r.h.s, if we can make sense of the derivative of ∂_{x^i} in the direction of $\dot{\gamma}$. This is exactly what the *covariant derivative* $\nabla_X Y$ does. If $Y \in T_m M$ then we can write $Y = a^i \partial_{x^i}$, and therefore

$$\nabla_X Y = \mathcal{L}_X a^i \partial_{x^i}. \quad (3.127)$$

Since there are several ways of choosing these coordinates, one must check that the definition does not depend on the choice. Note that for two vector-fields we define $(\nabla_Y X)(m) = \nabla_{Y(m)} X$. In the end we get a *connection*

$$\nabla : \mathcal{X}^k(M) \times \mathcal{X}^k(M) \rightarrow \mathcal{X}^k(M),$$

which satisfies (for all $f \in C^k(M, \mathbb{R})$ and $X, Y, Z \in \mathcal{X}^k(M)$):

- (1) $Y \rightarrow \nabla_Y X$ is tensorial, i.e., linear and $\nabla_{fY} X = f \nabla_Y X$.
- (2) $X \rightarrow \nabla_Y X$ is linear.
- (3) $\nabla_X(fY) = (\nabla_X f)Y + f \nabla_X Y$.
- (4) $\nabla_X Y - \nabla_Y X = [X, Y]$.
- (5) $\mathcal{L}_X g(Z, Y) = g(\nabla_X Z, Y) + g(Z, \nabla_X Y)$.

A semicolon is commonly used to denote covariant differentiation with respect to a natural basis vector. If $X = \partial_{x^i}$, then the components of $\nabla_X Y$ in (3.127) are denoted

$$Y_{;i}^k = \partial_{x^i} Y^k + \Gamma_{ij}^k Y^j, \quad (3.128)$$

where Γ_{ij}^k are *Christoffel symbols* defined in (3.129) below. Similar relations hold for higher-order tensor-fields (with as many terms with Christoffel symbols as is the tensor valence).

Therefore, no matter which coordinates we use, we can now define the acceleration of a curve in the following way:

$$\begin{aligned} \gamma(t) &= (\gamma^1(t), \dots, \gamma^n(t)), \\ \dot{\gamma}(t) &= \dot{\gamma}^i(t) \partial_{x^i}, \\ \ddot{\gamma}(t) &= \ddot{\gamma}^i(t) \partial_{x^i} + \dot{\gamma}^i(t) \nabla_{\dot{\gamma}(t)} \partial_{x^i}. \end{aligned}$$

We call γ a *geodesic* if $\gamma(t) = 0$. This is a second-order nonlinear ODE in a fixed coordinate system (x^1, \dots, x^n) at the specified point $m \in M$. Thus we see that given any tangent vector $X \in T_m M$, there is a unique geodesic $\gamma_X(t)$ with $\dot{\gamma}_X(0) = X$. If the manifold M is closed, the geodesic must exist for all time, but in case the manifold M is open this might not be so. To see this, take as M any open subset of Euclidean space with the induced metric.

Given an arbitrary vector-field $Y(t)$ along γ , i.e., $Y(t) \in T_{\gamma(t)} M$ for all t , we can also define the derivative $\dot{Y} \equiv \frac{dY}{dt}$ in the direction of $\dot{\gamma}$ by writing

$$\begin{aligned} Y(t) &= a^i(t) \partial_{x^i}, \\ \dot{Y}(t) &= \dot{a}^i(t) \partial_{x^i} + a^i(t) \nabla_{\dot{\gamma}(t)} \partial_{x^i}. \end{aligned}$$

Here the derivative of the tangent field $\dot{\gamma}$ is the acceleration γ . The field Y is said to be *parallel* iff $\dot{Y} = 0$. The equation for a field to be parallel is a first-order linear ODE, so we see that for any $X \in T_{\gamma(t_0)} M$ there is a unique parallel field $Y(t)$ defined on the entire domain of γ with the property that

$Y(t_0) = X$. Given two such parallel fields $Y, Z \in \mathcal{X}^k(M)$, we have that

$$\dot{g}(Y, Z) = D_{\dot{\gamma}}g(Y, Z) = g(\dot{Y}, Z) + g(Y, \dot{Z}) = 0.$$

Thus X and Y are both of constant length and form constant angles along γ . Hence, ‘parallel translation’ along a curve defines an orthogonal transformation between the tangent spaces to the manifold along the curve. However, in contrast to Euclidean space, this parallel translation will depend on the choice of curve.

An infinitesimal distance between the two nearby local points m and n on M is defined by an *arc-element*

$$ds^2 = g_{ij} dx^i dx^j,$$

and realized by the curves $x^i(s)$ of shortest distance, called *geodesics*, addressed by the *Hilbert 4th problem*. In local coordinates $(x^1(s), \dots, x^n(s))$ at a point $m \in M$, the geodesic defining equation is a second-order ODE,

$$\ddot{x}^i + \Gamma_{jk}^i \dot{x}^j \dot{x}^k = 0,$$

where the overdot denotes the derivative with respect to the affine parameter s , $\dot{x}^i(s) = \frac{d}{ds}x^i(s)$ is the tangent vector to the base geodesic, while the *Christoffel symbols* $\Gamma_{jk}^i = \Gamma_{jk}^i(m)$ of the *affine Levi-Civita connection* ∇ at the point $m \in M$ are defined, in a holonomic coordinate basis e_i as

$$\begin{aligned} \Gamma_{ij}^k &= g^{kl} \Gamma_{ijl}, \quad \text{with} \quad g^{ij} = (g_{ij})^{-1} \quad \text{and} \quad (3.129) \\ \Gamma_{ijk} &= \frac{1}{2} (\partial_{x^k} g_{ij} + \partial_{x^j} g_{ki} - \partial_{x^i} g_{jk}). \end{aligned}$$

Note that the Christoffel symbols (3.129) do not transform as tensors on the tangent bundle. They are the components of an object on the second tangent bundle, a *spray*. However, they do transform as tensors on the *jet space* (see section 5.3 below).

In *nonholonomic coordinates*, (3.129) takes the extended form

$$\Gamma_{kl}^i = \frac{1}{2} g^{im} (\partial_{x^l} g_{mk} + \partial_{x^k} g_{ml} - \partial_{x^m} g_{kl} + c_{mkl} + c_{mlk} - c_{klm}),$$

where $c_{klm} = g_{mp} c_{kl}^p$ are the commutation coefficients of the basis, i.e., $[e_k, e_l] = c_{kl}^m e_m$.

The *torsion* tensor-field T of the connection ∇ is the function $T : \mathcal{X}^k(M) \times \mathcal{X}^k(M) \rightarrow \mathcal{X}^k(M)$ given by

$$T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y].$$

From the skew symmetry ($[X, Y] = -[Y, X]$) of the Lie bracket, follows the skew symmetry ($T(X, Y) = -T(Y, X)$) of the torsion tensor. The mapping T is said to be f -bilinear since it is linear in both arguments and also satisfies $T(fX, Y) = fT(X, Y)$ for smooth functions f . Since $[\partial_{x^i}, \partial_{x^j}] = 0$ for all $1 \leq i, j \leq n$, it follows that

$$T(\partial_{x^i}, \partial_{x^j}) = (\Gamma_{ij}^k - \Gamma_{ji}^k) \partial_{x^k}.$$

Consequently, torsion T is a $(1, 2)$ tensor-field, locally given by

$$T = T_{ij}^k dx^i \otimes \partial_{x^k} \otimes dx^j,$$

where the torsion components T_{ij}^k are given by

$$T_{ij}^k = \Gamma_{ij}^k - \Gamma_{ji}^k.$$

Therefore, the torsion tensor gives a measure of the nonsymmetry of the connection coefficients. Hence, $T = 0$ if and only if these coefficients are symmetric in their subscripts. A connection ∇ with $T = 0$ is said to be *torsion free* or *symmetric*.

The connection also enables us to define many other classical concepts from calculus in the setting of Riemannian manifolds. Suppose we have a function $f \in C^k(M, \mathbb{R})$. If the manifold is not equipped with a Riemannian metric, then we have the differential of f defined by $df(X) = \mathcal{L}_X f$, which is a 1-form. The dual concept, the *gradient* of f , is supposed to be a vector-field. But we need a metric g to define it. Namely, ∇f is defined by the relationship

$$g(\nabla f, X) = df(X).$$

Having defined the gradient of a function on a Riemannian manifold, we can then use the connection to define the *Hessian* as the linear map

$$\nabla^2 f : TM \rightarrow TM, \quad \nabla^2 f(X) = \nabla_X \nabla f.$$

The corresponding bilinear map is then defined as

$$\nabla^2 f(X, Y) = g(\nabla^2 f(X), Y).$$

One can check that this is a symmetric bilinear form. The *Laplacian* of f , Δf , is now defined as the trace of the Hessian

$$\Delta f = \text{Tr}(\nabla^2 f(X)) = \text{Tr}(\nabla_X \nabla f),$$

which is a linear map. It is also called the *Laplace–Beltrami operator*, since Beltrami first considered this operator on Riemannian manifolds.

Riemannian metric has the following mechanical interpretation. Let M be a closed Riemannian manifold with the *mechanical metric* $g = g_{ij}v^i v^j \equiv \langle v, v \rangle$, with $v^i = \dot{x}^i$. Consider the *Lagrangian function*

$$L : TM \rightarrow \mathbb{R}, \quad (x, v) \mapsto \frac{1}{2} \langle v, v \rangle - U(x) \quad (3.130)$$

where $U(x)$ is a smooth function on M called the *potential*. On a fixed level of energy E , bigger than the maximum of U , the Lagrangian flow generated by (3.130) is conjugate to the geodesic flow with metric $\bar{g} = 2(e - U(x))\langle v, v \rangle$. Moreover, the reduced action of the Lagrangian is the distance for $g = \langle v, v \rangle$ [Arnold (1989); Abraham *et al.* (1988)]. Both of these statements are known as the *Maupertius action principle*.

3.10.1.2 Geodesics on M

For a $C^k, k \geq 2$ curve $\gamma : I \rightarrow M$, we define its *length* on I as

$$L(\gamma, I) = \int_I |\dot{\gamma}| \, dt = \int_I \sqrt{g(\dot{\gamma}, \dot{\gamma})} \, dt.$$

This length is independent of our parametrization of the curve γ . Thus the curve γ can be reparameterized, in such a way that it has unit velocity. The *distance between two points* m_1 and m_2 on M , $d(m_1, m_2)$, can now be defined as the infimum of the lengths of all curves from m_1 to m_2 , i.e.,

$$L(\gamma, I) \rightarrow \min.$$

This means that the distance measures the shortest way one can travel from m_1 to m_2 .

If we take a variation $V(s, t) : (-\varepsilon, \varepsilon) \times [0, \ell] \rightarrow M$ of a smooth curve $\gamma(t) = V(0, t)$ parameterized by arc-length L and of length ℓ , then the first derivative of the arc-length function

$$L(s) = \int_0^\ell |\dot{V}| \, dt, \quad \text{is given by}$$

$$\frac{dL(0)}{ds} \equiv \dot{L}(0) = g(\dot{\gamma}, X)|_0^\ell - \int_0^\ell g(\gamma, X) \, dt, \quad (3.131)$$

where $X(t) = \frac{\partial V}{\partial s}(0, t)$ is the so-called *variation vector-field*. Equation (3.131) is called the *first variation formula*. Given any vector-field X along γ , one can produce a variation whose variational field is X . If the variation

fixes the endpoints, $X(a) = X(b) = 0$, then the second term in the formula drops out, and we note that the length of γ can always be decreased as long as the acceleration of γ is not everywhere zero. Thus the *Euler–Lagrangian equations* for the arc-length functional are the equations for a curve to be a *geodesic*.

Recall that in local coordinates $x^i \in U$, where U is an open subset in the Riemannian manifold M , the geodesics are defined by the *geodesic equation*

$$\ddot{x}^i + \Gamma_{jk}^i \dot{x}^j \dot{x}^k = 0, \quad (3.132)$$

where overdot means derivative upon the line parameter s , while Γ_{jk}^i are Christoffel symbols of the affine Levi–Civita connection ∇ on M . From (3.132) it follows that the linear *connection homotopy*,

$$\bar{\Gamma}_{jk}^i = s\Gamma_{jk}^i + (1-s)\Gamma_{jk}^i, \quad (0 \leq s \leq 1),$$

determines the same geodesics as the original Γ_{jk}^i .

3.10.1.3 Riemannian Curvature on M

The *Riemann curvature tensor* is a rather ominous tensor of type $(1, 3)$; i.e., it has three vector variables and its value is a vector as well. It is defined through the Lie bracket (3.7.2) as

$$R(X, Y)Z = (\nabla_{[X, Y]} - [\nabla_X, \nabla_Y])Z = \nabla_{[X, Y]}Z - \nabla_X \nabla_Y Z + \nabla_Y \nabla_X Z.$$

This turns out to be a vector valued $(1, 3)$ –tensor–field in the three variables $X, Y, Z \in \mathcal{X}^k(M)$. We can then create a $(0, 4)$ –tensor,

$$R(X, Y, Z, W) = g(\nabla_{[X, Y]}Z - \nabla_X \nabla_Y Z + \nabla_Y \nabla_X Z, W).$$

Clearly this tensor is skew-symmetric in X and Y , and also in Z and $W \in \mathcal{X}^k(M)$. This was already known to Riemann, but there are some further, more subtle properties that were discovered a little later by Bianchi. The *Bianchi symmetry condition* reads

$$R(X, Y, Z, W) = R(Z, W, X, Y).$$

Thus the Riemann curvature tensor is a symmetric *curvature operator*

$$\mathfrak{R} : \Lambda^2 TM \rightarrow \Lambda^2 TM.$$

The *Ricci tensor* is the $(1, 1)$ – or $(0, 2)$ –tensor defined by

$$\text{Ric}(X) = R(\partial_{x^i}, X)\partial_{x^i}, \quad \text{Ric}(X, Y) = g(R(\partial_{x^i}, X)\partial_{x^i}, Y),$$

for any orthonormal basis (∂_{x^i}) . In other words, the Ricci curvature is a trace of the curvature tensor. Similarly one can define the *scalar curvature* as the trace

$$\text{scal}(m) = \text{Tr}(\text{Ric}) = \text{Ric}(\partial_{x^i}, \partial_{x^i}).$$

When the Riemannian manifold has dimension 2, all of these curvatures are essentially the same. Since $\dim \Lambda^2 TM = 1$ and is spanned by $X \wedge Y$ where $X, Y \in \mathcal{X}^k(M)$ form an orthonormal basis for $T_m M$, we see that the curvature tensor depends only on the scalar value

$$K(m) = R(X, Y, X, Y),$$

which also turns out to be the *Gaussian curvature*. The Ricci tensor is a *homothety*

$$\text{Ric}(X) = K(m)X, \quad \text{Ric}(Y) = K(m)Y,$$

and the scalar curvature is twice the Gauss curvature. In dimension 3 there are also some redundancies as $\dim TM = \dim \Lambda^2 TM = 3$. In particular, the Ricci tensor and the curvature tensor contain the same amount of information.

The *sectional curvature* is a kind of generalization of the Gauss curvature whose importance Riemann was already aware of. Given a 2-plane $\pi \subset T_m M$ spanned by an orthonormal basis $X, Y \in \mathcal{X}^k(M)$ it is defined as

$$\text{sec}(\pi) = R(X, Y, X, Y).$$

The remarkable observation by Riemann was that the *curvature operator* is a *homothety*, i.e., looks like $\mathfrak{R} = kI$ on $\Lambda^2 T_m M$ iff all sectional curvatures of planes in $T_m M$ are equal to k . This result is not completely trivial, as the sectional curvature is not the entire quadratic form associated to the symmetric operator \mathfrak{R} . In fact, it is not true that $\text{sec} \geq 0$ implies that the curvature operator is nonnegative in the sense that all its eigenvalues are nonnegative. What Riemann did was to show that our special coordinates (x^1, \dots, x^n) at m can be chosen to be *normal* at m , i.e., satisfy the condition

$$x^i = \delta_j^i x^j, \quad (\delta_j^i x^j = g_{ij})$$

on a neighborhood of m . One can show that such coordinates are actually exponential coordinates together with a choice of an orthonormal basis for

$T_m M$ so as to identify $T_m M$ with \mathbb{R}^n . In these coordinates one can then expand the metric as follows:

$$g_{ij} = \delta_{ij} - \frac{1}{3} R_{ikjl} x^k x^l + O(r^3).$$

Now the equations $x^i = g_{ij} x^j$ evidently give conditions on the curvatures R_{ijkl} at m .

If $\Gamma_{jk}^i(m) = 0$, the manifold M is flat at the point m . This means that the $(1, 3)$ curvature tensor, defined locally at $m \in M$ as

$$R_{ijk}^l = \partial_{x^j} \Gamma_{ik}^l - \partial_{x^k} \Gamma_{ij}^l + \Gamma_{rj}^l \Gamma_{ik}^r - \Gamma_{rk}^l \Gamma_{ij}^r,$$

also vanishes at that point, i.e., $R_{ijk}^l(m) = 0$.

Now, the rate of change of a vector-field A^k on the manifold M along the curve $x^i(s)$ is properly defined by the *absolute covariant derivative*

$$\frac{D}{ds} A^k = \dot{x}^i \nabla_i A^k = \dot{x}^i (\partial_{x^i} A^k + \Gamma_{ij}^k A^j) = \dot{A}^k + \Gamma_{ij}^k \dot{x}^i A^j.$$

By applying this result to itself, we can get an expression for the second covariant derivative of the vector-field A^k along the curve $x^i(s)$:

$$\frac{D^2}{ds^2} A^k = \frac{d}{ds} \left(\dot{A}^k + \Gamma_{ij}^k \dot{x}^i A^j \right) + \Gamma_{ij}^k \dot{x}^i (\dot{A}^j + \Gamma_{mn}^j \dot{x}^m A^n).$$

In the local coordinates $(x^1(s), \dots, x^n(s))$ at a point $m \in M$, if $\delta x^i = \delta x^i(s)$ denotes the *geodesic deviation*, i.e., the infinitesimal vector describing perpendicular separation between the two neighboring geodesics, passing through two neighboring points $m, n \in M$, then the *Jacobi equation of geodesic deviation* on the manifold M holds:

$$\frac{D^2 \delta x^i}{ds^2} + R_{jkl}^i \dot{x}^j \delta x^k \dot{x}^l = 0. \quad (3.133)$$

This equation describes the *relative acceleration* between two infinitesimally close facial geodesics, which is proportional to the facial curvature (measured by the Riemann tensor R_{jkl}^i at a point $m \in M$), and to the geodesic deviation δx^i . Solutions of equation (3.133) are called *Jacobi fields*.

In particular, if the manifold M is a 2D-surface in \mathbb{R}^3 , the Riemann curvature tensor simplifies into

$$R_{jmn}^i = \frac{1}{2} R g^{ik} (g_{km} g_{jn} - g_{kn} g_{jm}),$$

where R denotes the *scalar Gaussian curvature*. Consequently the equation of geodesic deviation (3.133) also simplifies into

$$\frac{D^2}{ds^2} \delta x^i + \frac{R}{2} \delta x^i - \frac{R}{2} \dot{x}^i (g_{jk} \dot{x}^j \delta x^k) = 0. \quad (3.134)$$

This simplifies even more if we work in a locally Cartesian coordinate system; in this case the covariant derivative $\frac{D^2}{Ds^2}$ reduces to an ordinary derivative $\frac{d^2}{ds^2}$ and the metric tensor g_{ij} reduces to identity matrix I_{ij} , so our 2D equation of geodesic deviation (3.134) reduces into a simple second-order ODE in just two coordinates x^i ($i = 1, 2$)

$$\ddot{x}^i + \frac{R}{2} \delta x^i - \frac{R}{2} \dot{x}^i (I_{jk} \dot{x}^j \delta x^k) = 0.$$

3.10.2 Global Riemannian Geometry

3.10.2.1 The Second Variation Formula

Cartan also establishes another important property of manifolds with non-positive curvature. First he observes that all spaces of constant zero curvature have torsion-free fundamental groups. This is because any isometry of finite order on Euclidean space must have a fixed point (the center of mass of any orbit is necessarily a fixed point). Then he notices that one can geometrically describe the L^∞ center of mass of finitely many points $\{m_1, \dots, m_k\}$ in Euclidean space as the unique minimum for the strictly convex function

$$x \rightarrow \max_{i=1, \dots, k} \frac{1}{2} \left\{ (d(m_i, x))^2 \right\}.$$

In other words, the center of mass is the center of the ball of smallest radius containing $\{m_1, \dots, m_k\}$. Now Cartan's observation from above was that the exponential map is expanding and globally distance nondecreasing as a map:

$$(T_m M, \text{Euclidean metric}) \rightarrow (T_m M, \text{with pull-back metric}).$$

Thus distance functions are convex in nonpositive curvature as well as in Euclidean space. Hence the above argument can in fact be used to conclude that any Riemannian manifold of nonpositive curvature must also have torsion free fundamental group.

Now, let us set up the *second variation formula* and explain how it is used. We have already seen the first variation formula and how it can

be used to characterize geodesics. Now suppose that we have a unit speed geodesic $\gamma(t)$ parameterized on $[0, \ell]$ and consider a variation $V(s, t)$, where $V(0, t) = \gamma(t)$. Synge then shows that $(\ddot{L} \equiv \frac{d^2 L}{ds^2})$

$$\ddot{L}(0) = \int_0^\ell \{g(\dot{X}, \dot{X}) - (g(\dot{X}, \dot{\gamma}))^2 - g(R(X, \dot{\gamma})X, \dot{\gamma})\} dt + g(\dot{\gamma}, A)|_0^\ell,$$

where $X(t) = \frac{\partial V}{\partial s}(0, t)$ is the variational vector-field, $\dot{X} = \nabla_{\dot{\gamma}} X$, and $A(t) = \nabla_{\frac{\partial V}{\partial s}} X$. In the special case where the variation fixes the endpoints, i.e., $s \rightarrow V(s, a)$ and $s \rightarrow V(s, b)$ are constant, the term with A in it falls out. We can also assume that the variation is perpendicular to the geodesic and then drop the term $g(\dot{X}, \dot{\gamma})$. Thus, we arrive at the following simple form:

$$\ddot{L}(0) = \int_0^\ell \{g(\dot{X}, \dot{X}) - g(R(X, \dot{\gamma})X, \dot{\gamma})\} dt = \int_0^\ell \{|\dot{X}|^2 - \sec(\dot{\gamma}, X)|X|^2\} dt.$$

Therefore, if the sectional curvature is nonpositive, we immediately observe that any geodesic locally minimizes length (that is, among close-by curves), even if it does not minimize globally (for instance γ could be a closed geodesic). On the other hand, in positive curvature we can see that if a geodesic is too long, then it cannot minimize even locally. The motivation for this result comes from the unit sphere, where we can consider geodesics of length $> \pi$. Globally, we know that it would be shorter to go in the opposite direction. However, if we consider a variation of γ where the variational field looks like $X = \sin(t \cdot \frac{\pi}{\ell}) E$ and E is a unit length parallel field along γ which is also perpendicular to γ , then we get

$$\begin{aligned} \ddot{L}(0) &= \int_0^\ell \left\{ |\dot{X}|^2 - \sec(\dot{\gamma}, X)|X|^2 \right\} dt \\ &= \int_0^\ell \left\{ \left(\frac{\pi}{\ell}\right)^2 \cdot \cos^2\left(t \cdot \frac{\pi}{\ell}\right) - \sec(\dot{\gamma}, X) \sin^2\left(t \cdot \frac{\pi}{\ell}\right) \right\} dt \\ &= \int_0^\ell \left(\left(\frac{\pi}{\ell}\right)^2 \cdot \cos^2\left(t \cdot \frac{\pi}{\ell}\right) - \sin^2\left(t \cdot \frac{\pi}{\ell}\right) \right) dt = -\frac{1}{2\ell} (\ell^2 - \pi^2), \end{aligned}$$

which is negative if the length ℓ of the geodesic is greater than π . Therefore, the variation gives a family of curves that are both close to and shorter than γ . In the general case, we can then observe that if $\sec \geq 1$, then for the same type of variation we get

$$\ddot{L}(0) \leq -\frac{1}{2\ell} (\ell^2 - \pi^2).$$

Thus we can conclude that, if the space is complete, then the diameter must be $\leq \pi$ because in this case any two points are joined by a segment, which cannot minimize if it has length $> \pi$. With some minor modifications one can now conclude that any complete Riemannian manifold (M, g) with $\sec \geq k^2 > 0$ must satisfy $\text{diam}(M, g) \leq \pi \cdot k^{-1}$. In particular, M must be compact. Since the universal covering of M satisfies the same curvature hypothesis, the conclusion must also hold for this space; hence M must have compact universal covering space and finite fundamental group.

In odd dimensions all spaces of constant positive curvature must be orientable, as orientation reversing orthogonal transformation on odd-dimensional spheres have fixed points. This can now be generalized to manifolds of varying positive curvature. Synge did it in the following way: Suppose M is not simply-connected (or not orientable), and use this to find a shortest closed geodesic in a free homotopy class of curves (that reverses orientation). Now consider parallel translation around this geodesic. As the tangent field to the geodesic is itself a parallel field, we see that parallel translation preserves the orthogonal complement to the geodesic. This complement is now odd dimensional (even dimensional), and by assumption parallel translation preserves (reverses) the orientation; thus it must have a fixed point. In other words, there must exist a closed parallel field X perpendicular to the closed geodesic γ . We can now use the above second variation formula

$$\ddot{L}(0) = \int_0^\ell \{|\dot{X}|^2 - |X|^2 \sec(\dot{\gamma}, X)\} dt + g(\dot{\gamma}, A)|_0^\ell = - \int_0^\ell |X|^2 \sec(\dot{\gamma}, X) dt.$$

Here the boundary term drops out because the variation closes up at the endpoints, and $\dot{X} = 0$ since we used a parallel field. In case the sectional curvature is always positive we then see that the above quantity is negative. But this means that the closed geodesic has nearby closed curves which are shorter. However, this is in contradiction with the fact that the geodesic was constructed as a length minimizing curve in a free homotopy class.

In 1941 Myers generalized the diameter bound to the situation where one only has a lower bound for the Ricci curvature. The idea is that $\text{Ric}(\dot{\gamma}, \dot{\gamma}) = \sum_{i=1}^{n-1} \sec(E_i, \dot{\gamma})$ for any set of vector-fields E_i along γ such that $\dot{\gamma}, E_1, \dots, E_{n-1}$ forms an orthonormal frame. Now assume that the fields are parallel and consider the $n-1$ variations coming from the variational vector-fields $\sin(t \cdot \frac{\pi}{\ell}) E_i$. Adding up the contributions from the

variational formula applied to these fields then induces

$$\begin{aligned}\sum_{i=1}^{n-1} \ddot{L}(0) &= \sum_{i=1}^{n-1} \int_0^\ell \left\{ \left(\frac{\pi}{\ell} \right)^2 \cdot \cos^2 \left(t \cdot \frac{\pi}{\ell} \right) - \sec(\dot{\gamma}, E_i) \sin^2 \left(t \cdot \frac{\pi}{\ell} \right) \right\} dt \\ &= \int_0^\ell \left\{ (n-1) \left(\frac{\pi}{\ell} \right)^2 \cdot \cos^2 \left(t \cdot \frac{\pi}{\ell} \right) - \operatorname{Ric}(\dot{\gamma}, \dot{\gamma}) \sin^2 \left(t \cdot \frac{\pi}{\ell} \right) \right\} dt.\end{aligned}$$

Therefore, if $\operatorname{Ric}(\dot{\gamma}, \dot{\gamma}) \geq (n-1)k^2$ (this is the Ricci curvature of S_k^n), then

$$\begin{aligned}\sum_{i=1}^{n-1} \ddot{L}(0) &\leq (n-1) \int_0^\ell \left\{ \left(\frac{\pi}{\ell} \right)^2 \cdot \cos^2 \left(t \cdot \frac{\pi}{\ell} \right) - k^2 \sin^2 \left(t \cdot \frac{\pi}{\ell} \right) \right\} dt \\ &= -(n-1) \frac{1}{2\ell} (\ell^2 k^2 - \pi^2),\end{aligned}$$

which is negative when $\ell > \pi \cdot k^{-1}$ (the diameter of S_k^n). Thus at least one of the contributions $\frac{d^2 L_i}{ds^2}(0)$ must be negative as well, implying that the geodesic cannot be a segment in this situation.

3.10.2.2 Gauss–Bonnet Formula

In 1926 Hopf proved that in fact there is a *Gauss–Bonnet formula* for all even-dimensional hypersurfaces $H^{2n} \subset \mathbb{R}^{2n+1}$. The idea is that the determinant of the differential of the Gauss map $G : H^{2n} \rightarrow S^{2n}$ is the Gaussian curvature of the hypersurface. Moreover, this is an intrinsically computable quantity. If we integrate this over the hypersurface, we get,

$$\frac{1}{\operatorname{Vol} S^{2n}} \int_H \det(DG) = \deg(G),$$

where $\deg(G)$ is the *Brouwer degree* of the *Gauss map*. Note that this can also be done for odd-dimensional surfaces, in particular curves, but in this case the degree of the Gauss map will depend on the embedding or immersion of the hypersurface. Instead one gets the so-called winding number. Hopf then showed, as Dyck had earlier done for surfaces, that $\deg(G)$ is always half the *Euler characteristic* of H , thus yielding

$$\frac{2}{\operatorname{Vol} S^{2n}} \int_H \det(DG) = \chi(H). \quad (3.135)$$

Since the l.h.s of this formula is in fact intrinsic, it is natural to conjecture that such a formula should hold for all manifolds.

3.10.2.3 Ricci Flow on M

Ricci flow, or the *parabolic Einstein equation*, was introduced by R. Hamilton in 1982 [Hamilton (1982)] in the form

$$\partial_t g_{ij} = -2R_{ij}. \quad (3.136)$$

Now, because of the minus sign in the front of the Ricci tensor R_{ij} in this equation, the solution metric g_{ij} to the Ricci flow shrinks in positive Ricci curvature direction while it expands in the negative Ricci curvature direction. For example, on the 2-sphere S^2 , any metric of positive Gaussian curvature will shrink to a point in finite time. Since the Ricci flow (3.136) does not preserve volume in general, one often considers the *normalized Ricci flow* defined by

$$\partial_t g_{ij} = -2R_{ij} + \frac{2}{n} r g_{ij}, \quad (3.137)$$

where $r = \int R dV / \int dV$ is the average scalar curvature. Under this normalized flow, which is equivalent to the (unnormalized) Ricci flow (3.136) by reparameterizing in time t and scaling the metric in space by a function of t , the volume of the solution metric is constant in time. Also that Einstein metrics (i.e., $R_{ij} = c g_{ij}$) are fixed points of (3.137).

Hamilton [Hamilton (1982)] showed that on a closed Riemannian 3-manifold M^3 with initial metric of positive Ricci curvature, the solution $g(t)$ to the normalized Ricci flow (3.137) exists for all time and the metrics $g(t)$ converge exponentially fast, as time t tends to the infinity, to a constant positive sectional curvature metric g_∞ on M^3 .

Since the Ricci flow lies in the realm of parabolic partial differential equations, where the prototype is the heat equation, here is a brief review of the *heat equation* [Cao and Chow (1999)].

Let (M^n, g) be a Riemannian manifold. Given a C^2 function $u : M \rightarrow \mathbb{R}$, its Laplacian is defined in local coordinates $\{x^i\}$ to be

$$\Delta u = \text{Tr}(\nabla^2 u) = g^{ij} \nabla_i \nabla_j u,$$

where $\nabla_i = \nabla_{\partial_{x^i}}$ is its associated covariant derivative (Levi-Civita connection). We say that a C^2 function $u : M^n \times [0, T] \rightarrow \mathbb{R}$, where $T \in (0, \infty]$, is a solution to the heat equation if

$$\partial_t u = \Delta u.$$

One of the most important properties satisfied by the heat equation is the maximum principle, which says that for any smooth solution to the heat equation, whatever pointwise bounds hold at $t = 0$ also hold for $t > 0$. Let $u : M^n \times [0, T) \rightarrow \mathbb{R}$ be a C^2 solution to the heat equation on a complete Riemannian manifold. If $C_1 \leq u(x, 0) \leq C_2$ for all $x \in M$, for some constants $C_1, C_2 \in \mathbb{R}$, then $C_1 \leq u(x, t) \leq C_2$ for all $x \in M$ and $t \in [0, T)$ [Cao and Chow (1999)].

Now, given a smooth manifold M , a one-parameter family of metrics $g(t)$, where $t \in [0, T)$ for some $T > 0$, is a solution to the Ricci flow if (3.136) is valid at all $x \in M$ and $t \in [0, T)$. The minus sign in the equation (3.136) makes the Ricci flow a *forward* heat equation [Cao and Chow (1999)] (with the normalization factor 2).

In local geodesic coordinates $\{x^i\}$, we have [Cao and Chow (1999)]

$$g_{ij}(x) = \delta_{ij} - \frac{1}{3} R_{ipjq} x^p x^q + O(|x|^3), \quad \text{therefore,} \quad \Delta g_{ij}(0) = -\frac{1}{3} R_{ij},$$

where Δ is the standard Euclidean Laplacian. Hence the Ricci flow is like the heat equation for a Riemannian metric

$$\partial_t g_{ij} = 6\Delta g_{ij}.$$

The practical study of the Ricci flow is made possible by the following short-time existence result: Given any smooth compact Riemannian manifold (M, g_o) , there exists a unique smooth solution $g(t)$ to the Ricci flow defined on some time interval $t \in [0, \epsilon)$ such that $g(0) = g_o$ [Cao and Chow (1999)].

Now, given that short-time existence holds for any smooth initial metric, one of the main problems concerning the Ricci flow is to determine under what conditions the solution to the normalized equation exists for all time and converges to a constant curvature metric. Results in this direction have been established under various curvature assumptions, most of them being some sort of positive curvature. Since the Ricci flow (3.136) does not preserve volume in general, one often considers, as we mentioned in the Introduction, the normalized Ricci flow (3.137). Under this flow, the volume of the solution $g(t)$ is independent of time.

To study the long-time existence of the normalized Ricci flow, it is important to know what kind of curvature conditions are preserved under the equation. In general, the Ricci flow tends to preserve some kind of positivity of curvatures. For example, positive scalar curvature is preserved in all dimensions. This follows from applying the maximum principle to the

evolution equation for scalar curvature R , which is

$$\partial_t R = \Delta R + 2 |R_{ij}|^2.$$

In dimension 3, positive Ricci curvature is preserved under the Ricci flow. This is a special feature of dimension 3 and is related to the fact that the Riemann curvature tensor may be recovered algebraically from the Ricci tensor and the metric in dimension 3. Positivity of sectional curvature is not preserved in general. However, the stronger condition of positive curvature operator is preserved under the Ricci flow.

3.10.2.4 Structure Equations on M

Let $\{X_a\}_{a=1}^m, \{Y_i\}_{i=1}^n$ be local orthonormal framings on M, N respectively and $\{e_i\}_{i=1}^n$ be the induced framing on E defined by $e_i = Y_i \circ \phi$, then there exist smooth local coframings $\{\omega_a\}_{a=1}^m, \{\eta_i\}_{i=1}^n$ and $\{\phi^* \eta_i\}_{i=1}^n$ on TM, TN and E respectively such that (locally)

$$g = \sum_{a=1}^m \omega_a^2 \quad \text{and} \quad h = \sum_{i=1}^n \eta_i^2.$$

The corresponding *first structure equations* are [Mustafa (1999)]:

$$\begin{aligned} d\omega_a &= \omega_b \wedge \omega_{ba}, & \omega_{ab} &= -\omega_{ba}, \\ d\eta_i &= \eta_j \wedge \eta_{ji}, & \eta_{ij} &= -\eta_{ji}, \\ d(\phi^* \eta_i) &= \phi^* \eta_j \wedge \phi^* \eta_{ji}, & \phi^* \eta_{ij} &= -\phi^* \eta_{ji}, \end{aligned}$$

where the unique 1-forms $\omega_{ab}, \eta_{ij}, \phi^* \eta_{ij}$ are the respective connection forms. The *second structure equations* are

$$\begin{aligned} d\omega_{ab} &= \omega_{ac} \wedge \omega_{cb} + \Omega_{ab}^M, & d\eta_{ij} &= \eta_{ik} \wedge \eta_{kj} + \Omega_{ij}^N, \\ d(\phi^* \eta_{ij}) &= \phi^* \eta_{ik} \wedge \phi^* \eta_{kj} + \phi^* \Omega_{ij}^N, \end{aligned}$$

where the curvature 2-forms are given by

$$\Omega_{ab}^M = -\frac{1}{2} R_{abcd}^M \omega_c \wedge \omega_d \quad \text{and} \quad \Omega_{ij}^N = -\frac{1}{2} R_{ijkl}^N \eta_k \wedge \eta_l.$$

The pull back map ϕ^* and the push forward map ϕ_* can be written as [Mustafa (1999)]

$$\phi^* \eta_i = f_{ia} \omega_a$$

for unique functions f_{ia} on $U \subset M$, so that

$$\phi_* = e_i \otimes \phi^* \eta_i = f_{ia} e_i \otimes \omega_a.$$

Note that ϕ_* is a section of the vector bundle $\phi^{-1}TN \otimes T^*M$.

The covariant differential operators are represented as

$$\nabla^M X_a = \omega_{ab} \otimes X_b, \quad \nabla^N Y_i = \eta_{ij} \otimes Y_j, \quad \nabla^* \omega_a = -\omega_{ca} \otimes \omega_c,$$

where ∇^* is the dual connection on the cotangent bundle T^*M .

Furthermore, the induced connection ∇^ϕ on E is

$$\nabla^\phi e_i = (\eta_{ij}(Y_k) \circ \phi) e_j \otimes f_{ka} \omega_a.$$

The components of the Ricci tensor and scalar curvature are defined respectively by

$$R_{ab}^M = R_{acbc}^M \quad \text{and} \quad R^M = R_{aa}^M.$$

Given a function $f : M \rightarrow \mathbb{R}$, there exist unique functions $f_{cb} = f_{bc}$ such that

$$df_c - f_b \omega_{cb} = f_{cb} \omega_b, \quad (3.138)$$

where $f_c = df(X_c)$ for a local orthonormal frame $\{X_c\}_{c=1}^m$. To prove this we take the exterior derivative of $df = \sum_{c=1}^m f_c \omega_c$ and using structure equations, we have

$$0 = [df_c \wedge \omega_c + f_{bc} \omega_b \wedge \omega_{bc}] = [(df_c - f_b \omega_{cb}) \wedge \omega_c].$$

Hence by *Cartan's lemma* (cf. [Willmore (1993)]), there exist unique functions $f_{cb} = f_{bc}$ such that

$$df_c - f_b \omega_{cb} = f_{cb} \omega_b.$$

The Laplacian of a function f on M is given by

$$\Delta f = -\text{Tr}(\nabla df),$$

that is, negative of the usual Laplacian on functions.

3.10.3 Application: Autonomous Lagrangian Dynamics

3.10.3.1 Basis of Lagrangian Dynamics

Recall that Riemannian metric $g = \langle, \rangle$ on the configuration manifold M is a positive-definite quadratic form $g : TM \rightarrow \mathbb{R}$, given in local coordinates $q^i \in U$ (U open in M) as

$$g_{ij} \mapsto g_{ij}(q, m) dq^i dq^j, \quad \text{where} \quad (3.139)$$

$$g_{ij}(q, m) = m_\mu \delta_{rs} \frac{\partial x^r}{\partial q^i} \frac{\partial x^s}{\partial q^j} \quad (3.140)$$

is the covariant material metric tensor defining a relation between internal and external coordinates and including n segmental masses m_μ . The quantities x^r are external coordinates ($r, s = 1, \dots, 6n$) and $i, j = 1, \dots, N \equiv 6n - h$, where h denotes the number of holonomic constraints.

The *Lagrangian* of the system is a quadratic form $L : TM \rightarrow \mathbb{R}$ dependent on velocity v and such that $L(v) = \frac{1}{2} \langle v, v \rangle$. It is locally given by

$$L(v) = \frac{1}{2} g_{ij}(q, m) v^i v^j.$$

On the velocity phase-space manifold TM exist:

- (1) a unique 1-form θ_L , defined in local coordinates $q^i, v^i = \dot{q}^i \in U_v$ (U_v open in TM) by $\theta_L = L_{v^i} dq^i$, where $L_{v^i} \equiv \partial L / \partial v^i$; and
- (2) a unique nondegenerate Lagrangian symplectic 2-form ω_L , which is closed ($d\omega_L = 0$) and exact ($\omega_L = d\theta_L = dL_{v^i} \wedge dq^i$).

TM is an orientable manifold, admitting the standard volume given by

$$\Omega_{\omega_L} = \frac{(-1)^{\frac{N(N+1)}{2}}}{N!} \omega_L^N,$$

in local coordinates $q^i, v^i = \dot{q}^i \in U_v$ (U_v open in TM) it is given by

$$\Omega_L = dq^1 \wedge \dots \wedge dq^N \wedge dv^1 \wedge \dots \wedge dv^N.$$

On the velocity phase-space manifold TM we can also define the *action* $A : TM \rightarrow \mathbb{R}$ in local coordinates $q^i, v^i = \dot{q}^i \in U_v$ (U_v open in TM) given by $A = v^i L_{v^i}$, so $E = v^i L_{v^i} - L$. The Lagrangian vector-field X_L on TM is determined by the condition $i_{X_L} \omega_L = dE$. Classically, it is given

by the second-order Lagrangian equations

$$\frac{d}{dt} \frac{\partial L}{\partial v^i} = \frac{\partial L}{\partial q^i}. \quad (3.141)$$

For a Lagrangian vector-field X_L on M , there is a base integral curve $\gamma_0(t) = (q^i(t), v^i(t))$ iff $\gamma_0(t)$ is a geodesic. This is given by the contravariant velocity equation

$$\dot{q}^i = v^i, \quad \dot{v}^i + \Gamma_{jk}^i v^j v^k = 0. \quad (3.142)$$

Here Γ_{jk}^i denote the *Christoffel symbols* of the Levi-Civita connection ∇ in an open chart U on M , defined on the Riemannian metric $g = \langle, \rangle$ by (see section 3.10.1.1 above)

$$\Gamma_{jk}^i = g^{il} \Gamma_{jkl}, \quad \Gamma_{ijk} = \frac{1}{2} (\partial_{x^i} g_{jk} + \partial_{x^j} g_{ki} + \partial_{x^k} g_{ij}). \quad (3.143)$$

The l.h.s $\dot{v}^i = \dot{v}^i + \Gamma_{jk}^i v^j v^k$ in the second part of (3.142) represents the *Bianchi covariant derivative* of the velocity with respect to t . *Parallel transport* on M is defined by $\dot{v}^i = 0$. When this applies, X_L is called the *geodesic spray* and its flow the *geodesic flow*.

For the dynamics in the gravitational potential field $V : M \rightarrow \mathbb{R}$, the Lagrangian $L : TM \rightarrow \mathbb{R}$ has an extended form

$$L(v, q) = \frac{1}{2} g_{ij} v^i v^j - V(q),$$

A Lagrangian vector-field X_L is still defined by the second-order Lagrangian equations (3.141, 3.142).

A general form of the forced, non-conservative Lagrangian equations is given as

$$\frac{d}{dt} \frac{\partial L}{\partial v^i} - \frac{\partial L}{\partial q^i} = F_i(t, q^i, v^i).$$

Here the $F_i(t, q^i, v^i)$ represent any kind of *covariant forces* as a functions of time, coordinates and momenta. In covariant form we have

$$\dot{q}^i = v^i, \quad g_{ij}(\dot{v}^i + \Gamma_{jk}^i v^j v^k) = F_j(t, q^i, v^i).$$

3.10.3.2 Lagrange–Poincaré Dynamics

Euler–Poincaré Equations

Let G be a Lie group and let $L : TG \rightarrow \mathbb{R}$ be a *left-invariant Lagrangian*. Let $l : \mathfrak{g} \rightarrow \mathbb{R}$ be its restriction to the identity. For a curve $g(t) \in G$, let $\xi(t) = g(t)^{-1} \cdot \dot{g}(t)$; that is, $\xi(t) = T_{g(t)}L_{g(t)^{-1}}\dot{g}(t)$. Then the following are equivalent [Marsden and Ratiu (1999)]:

- (1) $g(t)$ satisfies the Euler–Lagrangian equations for L on G ;
- (2) The variational principle holds,

$$\delta \int L(g(t), \dot{g}(t)) dt = 0$$

for variations with fixed endpoints;

- (3) The *Euler–Poincaré equations* hold:

$$\frac{d}{dt} \frac{\partial l}{\partial \xi} = \text{Ad}_\xi^* \frac{\delta l}{\delta \xi};$$

- (4) The variational principle holds on \mathfrak{g} ,

$$\delta \int l(\xi(t)) dt = 0,$$

using variations of the form $\delta \xi = \dot{\eta} + [\xi, \eta]$, where η vanishes at the endpoints.

Lagrange–Poincaré Equations

Here we follow [Marsden and Ratiu (1999)] and drop Euler–Lagrangian equations and variational principles from a general velocity phase-space TM to the quotient TM/G by an action of a Lie group G on M . If L is a G -invariant Lagrangian on TM , it induces a reduced Lagrangian l on TM/G . We introduce a connection A on the principal bundle $M \rightarrow S = M/G$, assuming that this quotient is nonsingular. This connection allows one to split the variables into a horizontal and vertical part. Let internal variables x^α be coordinates for shape-space $S = M/G$, let η^a be coordinates for the Lie algebra \mathfrak{g} relative to a chosen basis, let l be the Lagrangian regarded as a function of the variables $x^\alpha, \dot{x}^\alpha, \eta^a$ and let C_{db}^a be the structure constants of the Lie algebra \mathfrak{g} of G .

If one writes the Euler–Lagrangian equations on TM in a local principal bundle trivialization, with coordinates x^α on the base and η^a in the fibre,

then one gets the following system of *Hamel equations*:

$$\frac{d}{dt} \frac{\partial l}{\partial \dot{x}^\alpha} = \frac{\partial l}{\partial x^\alpha}, \quad \text{and} \quad \frac{d}{dt} \frac{\partial l}{\partial \eta^b} = \frac{\partial l}{\partial \eta^a} C_{db}^a \eta^a.$$

However, this representation of the equations does not make global intrinsic sense. The introduction of a connection overcomes this, and one can intrinsically and globally split the original variational principle relative to horizontal and vertical variations. One gets from one form to the other by means of the velocity shift given by replacing η^a by the vertical part relative to the *affine connection*

$$\xi^a = A_\alpha^a \dot{x}^\alpha + \eta^a.$$

Here A_α^a are the local coordinates of the connection A . This change of coordinates is motivated from the mechanical point of view, since the variables ξ^a have the interpretation of the locked angular velocity. The resulting *Lagrange–Poincaré equations* have the following form:

$$\begin{aligned} \frac{d}{dt} \frac{\partial l}{\partial \dot{x}^\alpha} - \frac{\partial l}{\partial x^\alpha} &= \frac{\partial l}{\partial \xi^a} \left(B_{\alpha\beta}^a \dot{x}^\beta + B_{\alpha d}^a \xi^d \right), \\ \frac{d}{dt} \frac{\partial l}{\partial \xi^b} &= \frac{\partial l}{\partial \xi^a} \left(B_{b\alpha}^a \dot{x}^\alpha + C_{db}^a \xi^d \right). \end{aligned}$$

In these equations, $B_{\alpha\beta}^a$ are the coordinates of the curvature B of A ,

$$B_{d\alpha}^a = C_{db}^a A_\alpha^b, \quad \text{and} \quad B_{b\alpha}^a = -B_{\alpha b}^a.$$

The variables ξ^a may be regarded as the rigid part of the variables on the original configuration space, while x^α are the internal variables.

3.10.4 Core Application: Search for Quantum Gravity

3.10.4.1 What Is Quantum Gravity?

The landscape of fundamental physics has changed substantially during the last few decades. Not long ago, our understanding of the weak and strong interactions was very confused, while general relativity was almost totally disconnected from the rest of physics and was empirically supported by little more than its three classical tests. Then two things have happened. The $SU(3) \times SU(2) \times U(1)$ *Standard Model* has found a dramatic empirical success, showing that *quantum field theory* (QFT) is capable of describing all accessible fundamental physics, or at least all non-gravitational physics. At the same time, *general relativity* (GR) has undergone an extraordinary

‘renaissance’, finding widespread application in astrophysics and cosmology, as well as novel vast experimental support – so that today GR is basic physics needed for describing a variety of physical systems we have access to, including advanced technological systems [Ashby (1997)].

These two parallel developments have moved fundamental physics to a position in which it has rarely been in the course of its history: We have today a group of fundamental laws, the Standard Model and GR, which –even if it cannot be regarded as a satisfactory global picture of Nature– is perhaps the best confirmed set of fundamental theories after Newton’s universal gravitation and Maxwell’s electromagnetism. More importantly, there aren’t today experimental facts that openly challenge or escape this set of fundamental laws. In this unprecedented state of affairs, a large number of theoretical physicists from different backgrounds have begun to address *the* piece of the puzzle which is clearly missing: combining the two halves of the picture and understanding the quantum properties of the gravitational field. Equivalently, understanding the quantum properties of space–time. Interest and researches in quantum gravity have thus increased sharply in recent years. And the problem of understanding what is a quantum space–time is today at the core of fundamental physics.

Today we have some well developed and reasonably well defined tentative theories of quantum gravity. String theory and loop quantum gravity are the two major examples. Within these theories definite physical results have been obtained, such as the explicit computation of the ‘quanta of geometry’ and the derivation of the black hole entropy formula. Furthermore, a number of fresh new ideas, like noncommutative geometry, have entered quantum gravity. For an overview of the problem of quantum gravity, see [Isham (1997)].

3.10.4.2 *Main Approaches to Quantum Gravity*

String theory

String theory is by far the research direction which is presently most investigated. String theory presently exists at two levels. First, there is a well developed set of techniques that define the string perturbation expansion over a given metric background. Second, the understanding of the non–perturbative aspects of the theory has much increased in recent years [Polchinski (1995)] and in the string community there is a widespread faith, supported by numerous indications, in the existence of a yet-to-be-found full non–perturbative theory, capable of generating the perturbation

expansion. There are attempts of constructing this non-perturbative theory, generically denoted M theory. The currently popular one is Matrix-theory, of which it is far too early to judge the effectiveness [Matacz (2002); Ishibashi *et. al.* (1997)].

The claim that string theory solves QG is based on two facts. First, the string perturbation expansion includes the graviton. More precisely, one of the string modes is a massless spin two, and helicity ± 2 , particle. Such a particle necessarily couples to the energy-momentum tensor of the rest of the fields [Weinberg (1964); Weinberg (1980)] and gives general relativity to a first approximation. Second, the perturbation expansion is consistent if the background geometry over which the theory is defined satisfies a certain consistency condition; this condition turns out to be a high energy modification of the Einstein's equations. The hope is that such a consistency condition for the perturbation expansion will emerge as a full-fledged dynamical equation from the yet-to-be-found non-perturbative theory.

From the point of view of the problem of quantum gravity, the relevant physical results from string theory are two [Rovelli (1997)]:

Black hole entropy. The most remarkable physical results for quantum gravity is the derivation of the *Bekenstein-Hawking formula* for the entropy of a black hole as a function of the horizon area. This beautiful result has been obtained by [Strominger and Vafa (1996)], and has then been extended in various directions. The result indicates that there is some unexpected internal consistency between string theory and QFT on curved space.

Microstructure of space-time. There are indications that in string theory the space-time continuum is meaningless below the Planck length. An old set of results on very high energy scattering amplitudes indicates that there is no way of probing the space-time geometry at very short distances. What happens is that in order to probe smaller distance one needs higher energy, but at high energy the string 'opens up from being a particle to being a true string' which is spread over space-time, and there is no way of focusing a string's collision within a small space-time region.

More recently, in the non-perturbative formulation of the Matrix-theory [Matacz (2002)], the space-time coordinates of the string x^i are replaced by matrices $(X^i)_m^n$. This can perhaps be viewed as a new interpretation of the space-time structure. The continuous space-time manifold emerges only in the long distance region, where these matrices are diagonal and com-

mute; while the space–time appears to have a noncommutative discretized structure in the short distance regime. These features are still poorly understood, but they have intriguing resonances with noncommutative geometry [Connes *et. al.* (1997)] and loop quantum gravity [Rovelli (1998)].

A key difficulty in string theory is the lack of a complete non-perturbative formulation. During the last year, there has been excitement for some tentative non-perturbative formulations [Matacz (2002)]; but it is far too early to understand if these attempts will be successful. Many previously highly acclaimed ideas have been rapidly forgotten.

A distinct and even more serious difficulty of string theory is the lack of a *background independent* formulation of the theory. In the words of Ed Witten:

‘Finding the right framework for an intrinsic, background independent formulation of string theory is one of the main problems in string theory, and so far has remained out of reach... This problem is fundamental because it is here that one really has to address the question of what kind of geometrical object the string represents.’

Most of string theory is conceived in terms of a theory describing excitations over this or that background, possibly with connections between different backgrounds. This is also true for (most) non-perturbative formulations such as Matrix theory. For instance, the (bosonic part of the) Lagrangian of Matrix-theory is

$$L \sim \frac{1}{2} \text{Tr} \left(\dot{X}^2 + \frac{1}{2} [X^i, X^j]^2 \right). \quad (3.144)$$

The indices that label the matrices X^i are raised and lowered with a Minkowski metric, and the theory is Lorentz invariant. In other words, the Lagrangian is really

$$L \sim \frac{1}{2} \text{Tr} \left(g^{00} g^{ij} \dot{X}_i \dot{X}_j + \frac{1}{2} g^{ik} g^{jl} [X_i, X_j] [X_k, X_l] \right), \quad (3.145)$$

where g is the flat metric of the background. This shows that there is a non-dynamical metric, and an implicit flat background in the action of the theory.

However, the world is not formed by a fixed background over which things happen. The background itself is dynamical. In particular, for instance, the theory should contain quantum states that are quantum superpositions of different backgrounds – and presumably these states play

an essential role in the deep quantum gravitational regime, namely in situations such as the big bang or the final phase of black hole evaporation. The absence of a fixed background in nature (or active diffeomorphism invariance) is the key general lessons we have learned from gravitational theories [Rovelli (1997)].

There has been a burst of recent activity in an outgrowth of string theory denoted *string cosmology* by [Veneziano (1991)]. The aim of string cosmology is to extract physical consequences from string theory by applying it to the big bang. The idea is to start from a Minkowski flat universe; show that this is unstable and therefore will run away from the flat (false-vacuum) state. The evolution then leads to a cosmological model that starts off in an inflationary phase. This scenario is described using mini-superspace technology, in the context of the low energy theory that emerge as limit of string theory. Thus, first one freezes all the massive modes of the string, then one freezes all massless modes except the zero modes (the spatially constant ones), obtaining a finite dimensional theory, which can be quantized non-perturbatively.

Loop quantum gravity

The second most popular approach to quantum gravity, and the most popular among relativists, is loop quantum gravity [Rovelli (1998)]. Loop quantum gravity is presently the best developed alternative to string theory. Like strings, it is not far from a complete and consistent theory and it yields a corpus of definite physical predictions, testable in principle, on quantum space-time.

Loop quantum gravity, however, attacks the problem from the opposite direction than string theory. It is a non-perturbative and background independent theory to start with. In other words, it is deeply rooted into the conceptual revolution generated by general relativity. In fact, successes and problems of loop quantum gravity are complementary to successes and problems of strings. Loop quantum gravity is successful in providing a consistent mathematical and physical picture of non perturbative quantum space-time; but the connection to the low energy dynamics is not yet completely clear.

The general idea on which loop quantum gravity is based is the following. The core of quantum mechanics is not identified with the structure of (conventional) QFT, because conventional QFT presupposes a background *metric* space-time, and is therefore immediately in conflict with GR.

Rather, it is identified with the general structure common to all quantum systems. The core of GR is identified with the absence of a fixed observable background space–time structure, namely with active diffeomorphism invariance. Loop quantum gravity is thus a quantum theory in the conventional sense: a Hilbert space and a set of quantum (field) operators, with the requirement that its classical limit is GR with its conventional matter couplings. But it is not a QFT over a metric manifold. Rather, it is a ‘quantum field theory on a differentiable manifold’, respecting the manifold’s invariances and where only coordinate independent quantities are physical.

Technically, loop quantum gravity is based on two inputs [Rovelli (1998); Rovelli (1997)]:

- The formulation of classical GR based on the Ashtekar connection [Ashtekar (1986); Ashtekar (1987); Ashtekar (1991)]. The version of the connection now most popular is not the original complex one, but an evolution of the same, in which the connection is real.
- The choice of the holonomies of this connection, denoted *loop variables*, as basic variables for the quantum gravitational field [Rovelli and Smolin (1988)].

This second choice determines the peculiar kind of quantum theory being built. Physically, it corresponds to the assumption that excitations with support on a loop are normalizable states. This is the key technical assumption on which everything relies.

It is important to notice that this assumption fails in conventional 4D *Yang–Mills theory*, because loop-like excitations on a metric manifold are too singular: the field needs to be smeared in more dimensions [Rovelli (1997)]. Equivalently, the linear closure of the loop states is a ‘far too big’ non-separable state space. This fact is the major source of some particle physicists’s suspicion at loop quantum gravity. What makes GR different from 4D Yang–Mills theory, however, is non-perturbative diffeomorphism invariance. The gauge invariant states, in fact, are not localized at all – they are, pictorially speaking, smeared by the (gauge) diffeomorphism group all over the coordinates manifold. More precisely, factoring away the diffeomorphism group takes us down from the state space of the loop excitations, which is ‘too big’, to a separable physical state space of the right size. Thus, the consistency of the loop construction relies heavily on diffeomorphism invariance. In other words, the diff-invariant invariant loop states (more precisely, the diff-invariant spin network states) are not

physical excitations of a field *on* space–time. They are excitations *of* space–time itself.

Loop quantum gravity was briefly described by [Rovelli (1997)] as follows:

Definition of theory. The mathematical structure of the theory has been put on a very solid basis. Early difficulties have been overcome. In particular, there were three major problems in the theory: the lack of a well defined scalar product, the overcompleteness of the loop basis, and the difficulty of treating the reality conditions.

- The problem of the lack of a scalar product on the Hilbert space has been solved with the definition of a diffeomorphism invariant measure on a space of connections [Ashtekar and Lewandowski (1995)]. Later, it has also become clear that the same scalar product can be defined in a purely algebraic manner [DePietri and Rovelli (1996)]. The state space of the theory is therefore a genuine Hilbert space \mathcal{H} .
- The overcompleteness of the loop basis has been solved by the introduction of the spin network states [Rovelli and Smolin (1995)]. A spin network is a graph carrying labels (related to $SU(2)$ representations and called ‘colors’) on its links and its nodes. Each spin network defines a spin network state, and the spin network states form a (genuine, non-overcomplete) orthonormal basis in \mathcal{H} .
- The difficulties with the reality conditions have been circumvented by the use of the real formulation [Barbero (1994); Barbero (1995a); Barbero (1995b); Thiemann (1996)]. The kinematics of loop quantum gravity is now defined with a level of rigor characteristic of mathematical physics [Ashtekar and Isham (1992); Ashtekar *et. al.* (1995)] and the theory can be defined using various alternative techniques [DePietri and Rovelli (1996); DePietri (1997)].

Hamiltonian constraint. A rigorous definition version of the Hamiltonian constraint equation has been constructed. This is anomaly free, in the sense that the constraints algebra closes (but see later on). The Hamiltonian has the crucial properties of acting on nodes only, which implies that its action is naturally discrete and combinatorial [Rovelli and Smolin (1988); Rovelli and Smolin (1994)]. This fact is at the roots

of the existence of exact solutions [Rovelli and Smolin (1988)], and of the possible finiteness of the theory

Matter. The old hope that QFT divergences could be cured by QG has recently received an interesting corroboration. The matter part of the Hamiltonian constraint is well-defined without need of renormalization. Thus, a main possible stumbling block is over: infinities did not appear in a place where they could very well have appeared [Rovelli (1997)].

Black hole entropy. The first important physical result in loop quantum gravity is a computation of black hole entropy [Krasnov (1997); Rovelli (1996a); Rovelli (1996b)].

Quanta of geometry. A very exciting development in quantum gravity in the last years has been by the computations of the quanta of geometry. That is, the computation of the discrete eigenvalues of area and volume.

In quantum gravity, any quantity that depends on the metric becomes an operator. In particular, so do the area A of a given (physically defined) surface, or the volume V of a given (physically defined) spatial region. In loop quantum gravity, these operators can be written explicitly. They are mathematically well defined self-adjoint operators in the Hilbert space \mathcal{H} . We know from quantum mechanics that certain physical quantities are quantized, and that we can compute their discrete values by computing the eigenvalues of the corresponding operator. Therefore, if we can compute the eigenvalues of the area and volume operators, we have a physical prediction on the possible quantized values that these quantities can take, at the Planck scale. These eigenvalues have been computed in loop quantum gravity. Here is for instance the main sequence of the spectrum of the area

$$A_{\vec{j}} = 8\pi\gamma\hbar G \sum_i \sqrt{j_i(j_i + 1)}. \quad (3.146)$$

$\vec{j} = (j_1, \dots, j_n)$ is an n -tuple of half-integers, labeling the eigenvalues, G and \hbar are the Newton and Planck constants, and γ is a dimensionless free parameter, denoted the so-called *Immirzi parameter* [Immirzi (1997)], not determined by the theory. A similar result holds for the volume. The spectrum (3.146) has been rederived and completed using various different techniques [DePietri and Rovelli (1996)]. These spectra represent solid results of loop quantum gravity. Under certain additional assumptions on the behavior of area and volume operators in the presence of matter, these results can be interpreted as a corpus of detailed quantitative predictions on hypothetical Planck scale observations.

Besides its direct relevance, the quantization of the area and the volume is of interest because it provides a physical picture of quantum space–time. The states of the spin network basis are eigenstates of some area and volume operators. We can say that a spin network carries quanta of area along its links, and quanta of volume at its nodes. The magnitude of these quanta is determined by the coloring. For instance, the half-integers $j_1 \dots j_n$ in (3.146) are the coloring of the spin network’s links that cross the given surface. Thus, a quantum space–time can be decomposed in a basis of states that can be visualized as made by quanta of volume (the intersections) separated by quanta of area (the links). More precisely, we can view a spin network as sitting on the *dual* of a cellular decomposition of physical space. The nodes of the spin network sit in the center of the 3-cells, and their coloring determines the (quantized) 3-cell’s volume. The links of the spin network cut the faces of the cellular decomposition, and their color \vec{j} determine the (quantized) areas of these faces via equation (3.146).

3.10.4.3 Traditional Approaches to Quantum Gravity

Discrete Approaches

Discrete quantum gravity is the program of regularizing classical GR in terms of some lattice theory, quantize this lattice theory, and then study an appropriate continuum limit, as one may do in QCD. There are three main ways of discretizing GR.

Regge Calculus

Regge introduced the idea of triangulating space–time by means of a simplicial complex and using the lengths l_i of the links of the complex as gravitational variables [Regge (1961)]. The theory can then be quantized by integrating over the lengths l_i of the links. For a recent review and extensive references see [Williams and Tuckey (1992)]. More recent work has focused in problems such as the geometry of Regge superspace [Hartle *et. al.* (1997)] and choice of the integration measure.

Dynamical Triangulations

Alternatively, one can keep the length of the links fixed, and capture the geometry by means of the way in which the simplices are glued together, namely by the triangulation. The *Einstein–Hilbert action* of Eu-

clidean gravity is approximated by a simple function of the total number of simplices and links, and the theory can be quantized summing over distinct triangulations (for a detailed introduction, see [Ambjørn *et. al.* (1998)]). There are two coupling constants in the theory, roughly corresponding to the Newton and cosmological constants. These define a two dimensional space of theories. The theory has a nontrivial continuum limit if in this parameter space there is a critical point corresponding to a second order phase transition. The theory has phase transition and a critical point. The transition separates a phase with crumpled space-times from a phase with ‘elongated’ spaces which are effectively 2D, with characteristic of a *branched polymer* [Bakker and Smit (1995); Ambjørn *et. al.* (2001a)]. This polymer structure is surprisingly the same as the one that emerges from loop quantum gravity at short scale. Near the transition, the model appears to produce ‘classical’ S^4 space-times, and there is evidence for scaling, suggesting a continuum behavior.

State Sum Models

A third road for discretizing GR was opened by a celebrated paper by [Ponzano and Regge (1968)]. They started from a Regge discretization of 3D GR and introduced a second discretization, by posing the so-called *Ponzano–Regge ansatz* that the lengths l assigned to the links are discretized as well, in half-integers in Planck units

$$l = \hbar G j, \quad j = 0, \quad \frac{1}{2}, 1, \dots \quad (3.147)$$

(Planck length is $\hbar G$ in 3D.) The half integers j associated to the links are denoted ‘coloring’ of the triangulation. Coloring can be viewed as the assignment of a $SU(2)$ irreducible representation to each link of the Regge triangulation. The elementary cells of the triangulation are tetrahedra, which have six links, colored with six $SU(2)$ representations. $SU(2)$ representation theory naturally assigns a number to a sextuplet of representations: the Wigner $6-j$ symbol. Rather magically, the product over all tetrahedra of these $6-j$ symbols converges to (the real part of the exponent of) the Einstein–Hilbert action. Thus, Ponzano and Regge were led to propose a quantization of 3D GR based on the partition function

$$Z \sim \sum_{\text{coloring}} \prod_{\text{tetrahedra}} 6-j(\text{color of the tetrahedron}), \quad (3.148)$$

where we have neglected some coefficients for simplicity. They also provided arguments indicating that this sum is independent from the triangulation of the manifold.

The formula (3.148) is simple and elegant, and the idea has recently had many surprising and interesting developments. 3D GR was quantized as a topological field theory by Ed Witten in [Witten (1988c)] and using loop quantum gravity in [Ashtekar *et. al.* (1989)]. The *Ponzano–Regge quantization* based on equation (3.148) was shown to be essentially equivalent to the TQFT quantization in [Ooguri (1992a)], and to the loop quantum gravity in [Rovelli (1993)] (for an extensive discussion of 3D quantum gravity, see [Carlip and Nelson (1995)]).

It turns out that the Ponzano–Regge ansatz (3.147) can be *derived* from loop quantum gravity [Rovelli (1993)]. Indeed, (3.147) is the 2D version of the 3D formula (3.146), which gives the quantization of the area. Therefore, a key result of quantum gravity of the last years, namely the quantization of the geometry, derived in the loop formalism from a full fledged non-perturbative quantization of GR, was anticipated as an ansatz by the intuition of Ponzano and Regge.

Hawking’s Euclidean Quantum Gravity

Hawking’s Euclidean quantum gravity is the approach based on his formal sum over Euclidean geometries (i.e., an Euclidean path integral, see chapter 6 below)

$$Z \sim \mathcal{N} \int \mathcal{D}[g] e^{-\int d^4x \sqrt{g} R[g]}. \quad (3.149)$$

As far as we understand, Hawking and his close collaborators do not anymore view this approach as an attempt to directly define a fundamental theory. The integral is badly ill defined, and does not lead to any known viable perturbation expansion. However, the main ideas of this approach are still alive in several ways.

First, Hawking’s picture of quantum gravity as a sum-over-space-times, continues to provide a powerful intuitive reference point for most of the research related to quantum gravity. Indeed, many approaches can be seen as attempts to replace the ill-defined and non-renormalizable formal integral (3.149) with a well defined expression. The dynamical triangulation approach (see above) and the spin foam approach (see below) are examples of attempts to realize Hawking’s intuition. Influence of Euclidean quantum gravity can also be found in the *Atiyah axioms* for TQFT.

Second, this approach can be used as an approximate method for describing certain regimes of non-perturbative quantum space-time physics, even if the fundamental dynamics is given by a more complete theory. In this spirit, Hawking and collaborators have continued the investigation of phenomena such as, for instance, pair creation of black holes in a background de Sitter space-time.

Effective Perturbative Quantum Gravity

If we expand classical GR around, say, the Minkowski metric,

$$g_{\mu\nu}(x) = \eta_{\mu\nu} + h_{\mu\nu}(x),$$

and construct a conventional QFT for the field $h_{\mu\nu}(x)$, we get, as it is well known, a non renormalizable theory. A small but intriguing group of papers has recently appeared, based on the proposal of treating this perturbative theory seriously, as a respectable low energy effective theory by its own. This cannot solve the deep problem of understanding the world in general relativistic quantum terms. But it can still be used for studying quantum properties of space-time in some regimes. This view has been advocated in a convincing way by John Donoghue, who has developed effective field theory methods for extracting physics from non renormalizable quantum GR [Donoghue (1996)].

QFT in Curved Space-Time

Quantum field theory in curved space-time is by now a reasonably established theory (see, e.g., [Wald (1994); Birrel and Davies (1982); Fulling (1989)], predicting physical phenomena of remarkable interest such as particle creation, vacuum polarization effects and Hawking's black-hole radiance [Hawking (1975)]. To be sure, there is no direct nor indirect experimental observation of any of these phenomena, but the theory is quite credible as an approximate theory, and many theorists in different fields would probably agree that these predicted phenomena are likely to be real.

The most natural and general formulation of the theory is within the algebraic approach [Haag (1992)], in which the primary objects are the local observables and the states of interest may all be treated on equal footing (as positive linear functionals on the algebra of local observables), even if they do not belong to the same Hilbert space.

The great merit of QFT on curved space-time is that it has provided us with some very important lessons. The key lesson is that in general

one loses the notion of a single preferred quantum state that could be regarded as *the* ‘vacuum’; and that the concept of ‘particle’ becomes vague and/or observer-dependent in a gravitational context. In a gravitational context, *vacuum* and *particle* are necessarily ill defined or approximate concepts. It is perhaps regrettable that this important lesson has not been yet absorbed by many scientists working in fundamental theoretical physics [Rovelli (1997)].

3.10.4.4 *New Approaches to Quantum Gravity*

Noncommutative Geometry

Noncommutative geometry is a research program in mathematics and physics which has recently received wide attention and raised much excitement. The program is based on the idea that space–time may have a noncommutative structure at the Planck scale. A main driving force of this program is the radical, volcanic and extraordinary sequence of ideas of A. Connes [Connes (1994)]. Connes observes that what we know about the structure of space–time derives from our knowledge of the fundamental interactions: special relativity derives from a careful analysis of Maxwell theory; Newtonian space–time and general relativity, derived both from a careful analysis of the gravitational interaction. Recently, we have learned to describe weak and strong interactions in terms of the $SU(3) \times SU(2) \times U(1)$ Standard Model. Connes suggests that the Standard Model might hide information on the minute structure of space–time as well. By making the hypothesis that the Standard Model symmetries reflect the symmetry of a noncommutative microstructure of space–time, Connes and Lott are able to construct an exceptionally simple and beautiful version of the Standard Model itself, with the impressive result that the Higgs field appears automatically, as the components of the Yang–Mills connection in the internal ‘noncommutative’ direction [Connes and Lott (1990)]. The theory admits a natural extension in which the space–time metric, or the gravitational field, is dynamical, leading to GR [Chamseddine and Connes (1996)].

The key idea behind a non-commutative space–time is to use algebra instead of geometry in order to describe spaces. Consider a topological (Hausdorff) space M . Consider all continuous functions f on M . These form an algebra A , because they can be multiplied and summed, and the algebra is commutative. According to a celebrated result, due to Gel’fand, knowledge of the algebra A is equivalent to knowledge of the space M ,

i.e., M can be reconstructed from A . In particular, the points x of the manifold can be obtained as the 1D irreducible representations x of A , which are all of the form $x(f) = f(x)$. Thus, we can use the algebra of the functions, instead of using the space. In a sense, notices Connes, the algebra is more physical, because we never deal with space–time: we deal with fields, or coordinates, over space–time. But one can capture Riemannian geometry as well, algebraically. Consider the Hilbert space H formed by all the spinor fields on a given Riemannian (spin) manifold. Let D be the (curved) Dirac operator, acting on H . We can view A as an algebra of (multiplicative) operators on H . Now, from the triple (H, A, D) , which Connes calls ‘spectral triple’, one can reconstruct the Riemannian manifold. In particular, it is not difficult to see that the distance between two points x and y can be obtained from these data by

$$d(x, y) = \sup_{\{f \in A, \|D \cdot f\| < 1\}} |x(f) - y(f)|, \quad (3.150)$$

a beautiful surprising algebraic definition of distance. A non-commutative space–time is the idea of describing space–time by a spectral triple in which the algebra A is a non-commutative algebra.

Remarkably, the gravitational field is captured, together with the Yang–Mills field, and the Higgs fields, by a suitable *Dirac operator* D [Chamseddine and Connes (1996)], and the full action is given simply by the trace of a very simple function of the Dirac operator.

Even if we disregard noncommutativity and the Standard Model, the above construction represents an intriguing re–formulation of conventional GR, in which the geometry is described by the Dirac operator instead than the metric tensor. This formulation has been explored in [Landi (1998)], where it is noticed that the eigenvalues of the Dirac operator are diffeomorphism invariant functions of the geometry, and therefore represent true observables in Euclidean GR. Their Poisson bracket algebra can be explicitly computed in terms of the energy–momentum eigenspinors. Surprisingly, the Einstein equations turn out to be captured by the requirement that the energy momentum of the eigen–spinors scale linearly with the eigenvalues.

Variants of Connes’s version of the idea of non commutative geometry and noncommutative coordinates have been explored by many authors (see, e.g., [Doplicher *et. al.* (1994)]) and intriguing connections with string theory have been suggested [Connes *et. al.* (1997); Fröhlich and Gawedzki (1994)].

Null–Surface Formulation

A second new set of ideas comes from [Frittelli *et. al.* (1995)]. These authors have discovered that the (conformal) information about the geometry is captured by suitable families of null hypersurfaces in space–time, and have been able to reformulate GR as a theory of self–interacting families of surfaces. A remarkable aspect of the theory is that physical information about the space–time interior is transferred to null infinity, along null geodesics. Thus, the space–time interior is described in terms of how we would (literally) ‘see it’ from outside. This description is diffeomorphism invariant, and addresses directly the relational localization characteristic of GR: the space–time location of a region is determined dynamically by the gravitational field and is captured by when and where we see the space–time region from infinity. This idea may lead to interesting and physically relevant diffeomorphism invariant observables in quantum gravity. A discussion of the quantum gravitational fuzziness of the space–time points determined by this perspective can be found in [Frittelli *et. al.* (1997)].

Spin Foam Models

From the mathematical point of view, the problem of quantum gravity is to understand what is QFT on a differentiable manifold without metric. A class of well understood QFT’s on manifolds exists. These are the topological quantum field theories (TQFT). Topological field theories are particularly simple field theories. They have as many fields as gauges and therefore no local degree of freedom, but only a finite number of global degrees of freedom. An example is GR in 3D, say on a torus (the theory is equivalent to a Chern–Simons theory). In 3D, the Einstein equations require that the geometry is flat, so there are no gravitational waves. Nevertheless, a careful analysis reveals that the radii of the torus are dynamical variables, governed by the theory. Witten has noticed that theories of this kind give rise to interesting quantum models [Witten (1988a)], and [Atiyah (1989)] has provided a beautiful axiomatic definition of a TQFT. Concrete examples of TQFT have been constructed using Hamiltonian, combinatorial and path integral methods. The relevance of TQFT for quantum gravity has been suggested by many and the recent developments have confirmed these suggestions.

Recall that TQFT is a diffeomorphism invariant QFT. Sometimes, the expression TQFT is used to indicate all diffeomorphism invariant QFT’s. This has lead to a widespread, but incorrect belief that any diffeomorphism

invariant QFT has a finite number of degrees of freedom, unless the invariance is somehow broken, for instance dynamically. This belief is wrong. The problem of quantum gravity is precisely to define a diffeomorphism invariant QFT having an infinite number degrees of freedom and ‘local’ excitations. Locality in a gravity theory, however, is different from locality in conventional field theory. This point is often source of confusion. Here is Rovelli’s clarification [Rovelli (1997)]:

- In a conventional field theory on a metric space, the degrees of freedom are local in the sense that they can be localized on the metric manifold (an electromagnetic wave is here or there in Minkowski space).
- In a diffeomorphism invariant field theory such as general relativity, the degrees of freedom are still local (gravitational waves exist), but they are not localized with respect to the manifold. They are nevertheless localized with respect to each other (a gravity wave is three meters apart from another gravity wave, or from a black hole).
- In a topological field theory, the degrees of freedom are not localized at all: they are global, and in finite number (the radius of a torus is not in a particular position on the torus).

The first TQFT directly related to quantum gravity was defined by [Turaev and Viro (1992)]. The Turaev–Viro model is a mathematically rigorous version of the 3D Ponzano–Regge quantum gravity model described above. In the Turaev–Viro theory, the sum (3.148) is made finite by replacing $SU(2)$ with quantum $SU(2)_q$ (with a suitable q). Since $SU(2)_q$ has a finite number of irreducible representations, this trick, suggested by [Ooguri (1992a); Ooguri (1992b)], makes the sum finite. The extension of this model to four dimensions has been actively searched for a while and has finally been constructed by [Crane and Yetter (1993)], again following Ooguri’s ideas. The Crane–Yetter (CY) model is the first example of 4D TQFT. It is defined on a simplicial decomposition of the manifold. The variables are spins (‘colors’) attached to faces and tetrahedra of the simplicial complex. Each 4-simplex contains 10 faces and 5 tetrahedra, and therefore there are 15 spins associated to it. The action is defined in terms of the quantum Wigner $15 - j$ symbols, in the same manner in which the Ponzano–Regge action is constructed in terms of products of $6 - j$ symbols.

$$Z \sim \sum_{\text{coloring}} \prod_{4\text{-simplices}} 15 - j(\text{color of the } 4\text{-simplex}), \quad (3.151)$$

(where we have disregarded various factors for simplicity). Crane and Yet-

ter introduced their model independently from loop quantum gravity. However, recall that loop quantum gravity suggests that in 4 dimensions the naturally discrete geometrical quantities are area and volume, and that it is natural to extend the Ponzano–Regge model to 4D by assigning colors to faces and tetrahedra.

The CY model is not a quantization of 4D GR, nor could it be, being a TQFT in strict sense. Rather, it can be formally derived as a quantization of $SU(2)$ BF theory. BF theory is a topological field theory with two fields, a connection A , with curvature F , and a 2-form B [Horowitz (1989)], with action

$$S[A, B] = \int B \wedge F. \quad (3.152)$$

However, there is a strict relation between GR and BF. If we add to $SO(3, 1)$ BF theory the constraint that the 2-form B is the product of two tetrad 1-forms

$$B = E \wedge E, \quad (3.153)$$

we get precisely GR. This observation has lead many to suggest that a quantum theory of gravity could be constructed by a suitable modification of quantum BF theory [Baez (1996c)]. This suggestion has become very plausible, with the following construction of the *spin foam models*.

The key step in development of the spin foam models was taken by [Barbieri (1997)], studying the ‘quantum geometry’ of the simplices that play a role in loop quantum gravity. Barbieri discovered a simple relation between the quantum operators representing the areas of the faces of the tetrahedra. This relation turns out to be the quantum version of the constraint (3.153), which turns BF theory into GR. [Barret and Crane (1997)] added the Barbieri relation to (the $SO(3, 1)$ version of) the CY model. This is equivalent to replacing the the 15 - j Wigner symbol, with a different function A_{BC} of the colors of the 4-simplex. This replacement defines a ‘modified TQFT’, which has a chance of having general relativity as its classical limit.

The Barret–Crane model is not a TQFT in strict sense. In particular, it is not independent from the triangulation. Thus, a continuum theory has to be formally defined by some suitable sum over triangulations

$$Z \sim \sum_{\text{triang}} \sum_{\text{coloring}} \prod_{4\text{-simplices}} A_{BC}(\text{color of the } 4\text{-simplex}). \quad (3.154)$$

This essential aspect of the construction, however, is not yet understood.

The Barret Crane model can virtually be obtained also from loop quantum gravity. This is an unexpected convergence of two very different lines of research. Loop quantum gravity is formulated canonically in the frozen time formalism. While the frozen time formalism is in principle complete, in practice it is cumbersome, and anti-intuitive. Our intuition is four dimensional, not three dimensional. An old problem in loop quantum gravity has been to derive a space–time version of the theory. A space–time formulation of quantum mechanics is provided by the sum over histories. A sum over histories can be derived from the Hamiltonian formalism, as Feynman did originally. Loop quantum gravity provides a mathematically well defined Hamiltonian formalism, and one can therefore follow Feynman steps and construct a sum over histories quantum gravity starting from the loop formalism. This has been done in [Reisenberger and Rovelli (1997)]. The sum over histories turns out to have the form of a sum over surfaces.

More precisely, the transition amplitude between two spin network states turns out to be given by a sum of terms, where each term can be represented by a (2D) branched ‘colored’ surface in space–time. A branched colored surface is formed by elementary surface elements carrying a label, that meet on edges, also carrying a label; edges, in turn meet in vertices (or branching points, see Figure 3.11). The contribution of one such sur-

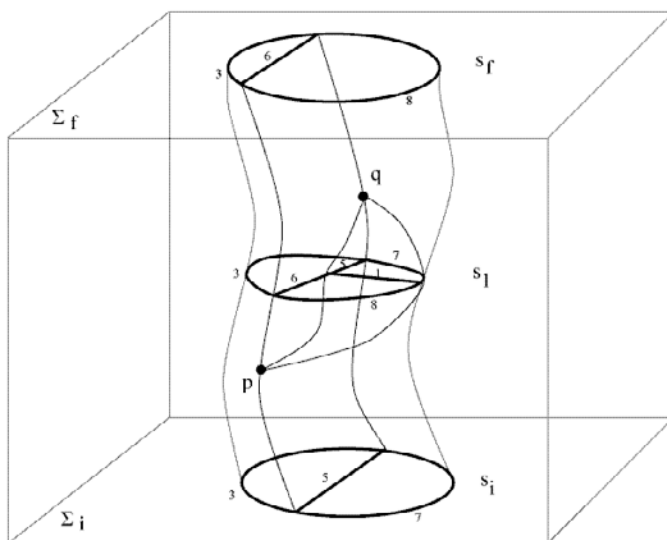


Fig. 3.11 A branched surface with two vertices.

faces to the sum over histories is the product of one term per each branching point of the surface. The branching points represent the ‘vertices’ of this theory, in the sense of Feynman. The contribution of each vertex can be computed algebraically from the ‘colors’ (half integers) of the adjacent surface elements and edges. Thus, space–time loop quantum gravity is defined by the partition function

$$Z \sim \sum_{\text{surfaces}} \sum_{\text{colorings}} \prod_{\text{vertices}} A_{\text{loop}}(\text{color of the vertex}) \quad (3.155)$$

The vertex A_{loop} is determined by a matrix elements of the Hamiltonian constraint. The fact that one obtains a sum over surfaces is not too surprising, since the time evolution of a loop is a surface. Indeed, the time evolution of a spin network (with colors on links and nodes) is a surface (with colors on surface elements and edges) and the Hamiltonian constraint generates branching points in the same manner in which conventional Hamiltonians generate the vertices of the Feynman diagrams.

Now, (3.155) has the same structure of the Barret–Crane model (3.151). To see this, simply notice that we can view each branched colored surface as located on the lattice *dual* to a triangulation. Then each vertex correspond to a 4-simplex; the coloring of the two models matches exactly (elementary surfaces \rightarrow faces, edges \rightarrow tetrahedra); and summing over surfaces corresponds to summing over triangulations. The main difference is the different weight at the vertices. The Barret–Crane vertex A_{BC} can be read as a covariant definition a Hamiltonian constraint in loop quantum gravity.

Thus, the space–time formulation of loop quantum GR is a simple modification of a TQFT. This approach provides a 4D pictorial intuition of quantum space–time, analogous to the Feynman graphs description of quantum field dynamics. John Baez has introduced the term ‘spin foam’ for the branched colored surfaces of the model, in honor of John Wheeler’s intuitions on the quantum microstructure of space–time. Spin foams are a precise mathematical implementation of Wheeler’s ‘space–time foam’ suggestions.

3.10.4.5 *Black Hole Entropy*

A focal point of the research in quantum gravity in the last years has been the discussion of black hole (BH) entropy. This problem has been discussed from a large variety of perspectives and within many different research programs.

Let us very briefly recall the origin of the problem. In *classical* GR, future event horizons behave in a manner that has a peculiar thermodynamical flavor. This remark, together with a detailed physical analysis of the behavior of hot matter in the vicinity of horizons, prompted Bekenstein to suggest that there is entropy associated to every horizon. The suggestion was first considered ridiculous, because it implies that a black hole is hot and radiates. But then Steven Hawking, in a celebrated work [Hawking (1975)], showed that QFT in curved space–time predicts that a black hole emits thermal radiation, precisely at the temperature predicted by Bekenstein, and Bekenstein’s courageous suggestion was fully vindicated. Since then, the entropy of a BH has been indirectly computed in a surprising variety of manners, to the point that BH entropy and BH radiance are now considered almost an established fact by the community, although, of course, they were never observed nor, presumably, they are going to be observed soon. This confidence, perhaps a bit surprising to outsiders, is related to the fact that thermodynamics is powerful in indicating general properties of systems, even if we do not control its microphysics. Many hope that the Bekenstein–Hawking radiation could play for quantum gravity a role analogous to the role played by the black body radiation for quantum mechanics. Thus, indirect arguments indicate that a Schwarzschild BH has an entropy

$$S = \frac{1}{4} \frac{A}{\hbar G} \quad (3.156)$$

The remaining challenge is to derive this formula from first principles [Rovelli (1997)].

Later in the book we will continue our exposition of various approaches to quantum gravity.

3.10.5 *Basics of Morse and (Co)Bordism Theories*

3.10.5.1 *Morse Theory on Smooth Manifolds*

At the same time the variational formulae were discovered, a related technique, called *Morse theory*, was introduced into Riemannian geometry. This theory was developed by Morse, first for functions on manifolds in 1925, and then in 1934, for the loop space. The latter theory, as we shall see, sets up a very nice connection between the first and second variation formulae from the previous section and the topology of M . It is this relationship that we shall explore at a general level here. In section 5 we shall then see how

this theory was applied in various specific settings.

If we have a *proper function* $f : M \rightarrow \mathbb{R}$, then its Hessian (as a quadratic form) is in fact well defined at its *critical points* without specifying an underlying Riemannian metric. The *nullity* of f at a *critical point* is defined as the dimension of the *kernel* of $\nabla^2 f$, while the *index* is the number of negative eigenvalues counted with multiplicity. A function is said to be a *Morse function* if the nullity at any of its critical points is zero. Note that this guarantees in particular that all critical points are isolated. The first fundamental Theorem of Morse theory is that one can determine the topological structure of a manifold from a Morse function. More specifically, if one can order the critical points x_1, \dots, x_k so that $f(x_1) < \dots < f(x_k)$ and the index of x_i is denoted λ_i , then M has the structure of a CW complex with a cell of dimension λ_i for each i . Note that in case M is closed then x_1 must be a minimum and so $\lambda_1 = 0$, while x_k is a maximum and $\lambda_k = n$. The classical example of Milnor of this Theorem in action is a torus in 3-space and f the height function.

We are now left with the problem of trying to find appropriate Morse functions. While there are always plenty of such functions, there does not seem to be a natural way of finding one. However, there are natural choices for Morse functions on the loop space to a Riemannian manifold. This is, somewhat inconveniently, infinite-dimensional. Still, one can develop Morse theory as above for suitable functions, and moreover *the loop space of a manifold determines the topology of the underlying manifold*.

If $m, p \in M$, then we denote by Ω_{mp} the space of all C^k paths from m to p . The first observation about this space is that

$$\pi_{i+1}(M) = \pi_i(\Omega_{mp}).$$

To see this, just fix a path from m to q and then join this path to every curve in Ω_{mp} . In this way Ω_{mp} is identified with Ω_m , *the space of loops fixed at m* . For this space the above relationship between the homotopy groups is almost self-evident.

On the space Ω_{mp} we have two naturally defined functions, the *arc-length* and *energy functionals*:

$$L(\gamma, I) = \int_I |\dot{\gamma}| \, dt, \quad \text{and} \quad E(\gamma, I) = \frac{1}{2} \int_I |\dot{\gamma}|^2 \, dt.$$

While the energy functional is easier to work with, it is the arc-length functional that we are really interested in. In order to make things work out nicely for the arc-length functional, it is convenient to parameterize

all curves on $[0, 1]$ and proportionally to arc-length. We shall think of Ω_{mp} as an *infinite-dimensional manifold*. For each curve $\gamma \in \Omega_{mp}$ the natural choice for the tangent space consists of the vector-fields along γ which vanish at the endpoints of γ . This is because these vector-fields are exactly the variational fields for curves through γ in Ω_{mp} , i.e., fixed endpoint variations of γ . An inner product on the tangent space is then naturally defined by

$$(X, Y) = \int_0^1 g(X, Y) dt.$$

Now the first variation formula for arc-length tells us that the gradient for L at γ is $-\nabla_{\dot{\gamma}}\dot{\gamma}$. Actually this cannot be quite right, as $-\nabla_{\dot{\gamma}}\dot{\gamma}$ does not vanish at the endpoints. The real gradient is gotten in the same way we find the gradient for a function on a surface in space, namely, by projecting it down into the correct tangent space. In any case we note that the critical points for L are exactly the geodesics from m to p . The second variation formula tells us that the Hessian of L at these critical points is given by

$$\nabla^2 L(X) = \ddot{X} + R(X, \dot{\gamma}) \dot{\gamma},$$

at least for vector-fields X which are perpendicular to γ . Again we ignore the fact that we have the same trouble with endpoint conditions as above. We now need to impose the Morse condition that this Hessian is not allowed to have any kernel. The vector-fields J for which $\ddot{J} + R(J, \dot{\gamma}) \dot{\gamma} = 0$ are called *Jacobi fields*. Thus we have to Figure out whether there are any Jacobi fields which vanish at the endpoints of γ . The first observation is that Jacobi fields must always come from geodesic variations. The Jacobi fields which vanish at m can therefore be found using the exponential map \exp_m . If the Jacobi field also has to vanish at p , then p must be a critical value for \exp_m . Now Sard's Theorem asserts that the set of critical values has measure zero. For given $m \in M$ it will therefore be true that the arc-length functional on Ω_{mp} is a Morse function for almost all $p \in M$. Note that it may not be possible to choose $p = m$, the simplest example being the standard sphere. We are now left with trying to decide what the *index* should be. This is the dimension of the largest subspace on which the Hessian is negative definite. It turns out that this index can also be computed using Jacobi fields and is in fact always finite. Thus one can calculate the topology of Ω_{mp} , and hence M , by finding all the geodesics from m to p and then computing their index.

In geometrical situations it is often unrealistic to suppose that one can calculate the index precisely, but as we shall see it is often possible to given lower bounds for the index. As an example, note that if M is not simply-connected, then Ω_{mp} is not connected. Each curve of minimal length in the path components is a geodesic from m to p which is a local minimum for the arc-length functional. Such geodesics evidently have index zero. In particular, if one can show that all geodesics, except for the minimal ones from m to p , have index > 0 , then the manifold must be simply-connected. We will apply Morse theory in biodynamics/robotic in section (3.13.5.2) below.

3.10.5.2 (Co)Bordism Theory on Smooth Manifolds

(Co)bordism appeared as a revival of Poincaré's unsuccessful 1895 attempts to define homology using only manifolds. Smooth manifolds (without boundary) are again considered as 'negligible' when they are *boundaries* of smooth manifolds-with-boundary. But there is a big difference, which keeps definition of 'addition' of manifolds from running into the difficulties encountered by Poincaré; it is now the disjoint union. The (unoriented) *(co)bordism relation* between two compact smooth manifolds M_1, M_2 of same dimension n means that their disjoint union $\partial W = M_1 \cup M_2$ is the boundary ∂W of an $(n+1)$ D smooth manifold-with-boundary W . This is an *equivalence relation*, and the classes for that relation of n D manifolds form a *commutative group* \mathfrak{N}_n in which every element has order 2. The direct sum $\mathfrak{N}_\bullet = \bigoplus_{n \geq 0} \mathfrak{N}_n$ is a ring for the multiplication of classes deduced from the Cartesian product of manifolds.

More precisely, a manifold M is said to be a (co)bordism from A to B if exists a diffeomorphism from a disjoint sum, $\varphi \in \text{diff}(A^* \cup B, \partial M)$. Two (co)bordisms $M(\varphi)$ and $M'(\varphi')$ are equivalent if there is a $\Phi \in \text{diff}(M, M')$ such that $\varphi' = \Phi \circ \varphi$. The equivalence class of (co)bordisms is denoted by $M(A, B) \in \text{Cob}(A, B)$ [Stong (1968)].

Composition c_{Cob} of (co)bordisms comes from gluing of manifolds [Baez and Dolan (1995)]. Let $\varphi' \in \text{diff}(C^* \cup D, \partial N)$. One can glue (co)bordism M with N by identifying B with C^* , $(\varphi')^{-1} \circ \varphi \in \text{diff}(B, C^*)$. We get the glued (co)bordism

$(M \circ N)(A, D)$ and a semigroup operation,

$$c(A, B, D) : \text{Cob}(A, B) \times \text{Cob}(B, D) \longrightarrow \text{Cob}(A, D).$$

A *surgery* is an operation of cutting a manifold M and gluing to cylin-

ders. A surgery gives new (co)bordism: from $M(A, B)$ into $N(A, B)$. The disjoint sum of $M(A, B)$ with $N(C, D)$ is a (co)bordism $(M \cup N)(A \cup C, B \cup D)$. We got a 2-graph of (co)bordism Cob with $Cob_0 = Man_d$, $Cob_1 = Man_{d+1}$, whose 2-cells from Cob_2 are surgery operations.

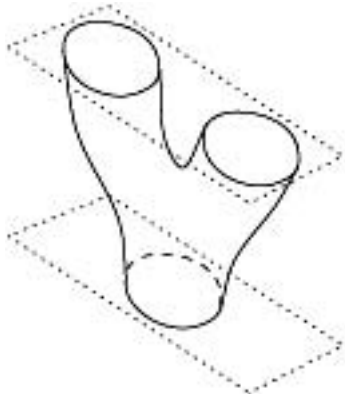
There is an n -category of (co)bordisms \mathcal{BO} [Leinster (2003)] with:

- 0-cells: 0-manifolds, where ‘manifold’ means ‘compact, smooth, oriented manifold’. A typical 0-cell is
- 1-cells: 1-manifolds with corners, i.e., (co)bordisms between



0-manifolds, such as (this being a 1-cell from the 4-point manifold to the 2-point 0-manifold).

- 2-cells: 2-manifolds with corners, such as



- 3-cells, 4-cells,... are defined similarly;
- Composition is gluing of manifolds.

The (co)bordisms theme was taken a step further by [Baez and Dolan (1995)], when when they started a programme to understand the subtle relations between certain TMFT models for manifolds of different dimensions, frequently referred to as the dimensional ladder. This programme is based on higher-dimensional algebra, a generalization of the theory of categories and functors to n -categories and n -functors. In this framework a topological quantum field theory (TMFT) becomes an n -functor from the n -category \mathcal{BO} of n -cobordisms to the n -category of n -Hilbert spaces.

3.11 Finsler Manifolds and Their Applications

Recall that Finsler geometry is such a generalization of Riemannian geometry, that is closely related to multivariable calculus of variations.

3.11.1 Definition of a Finsler Manifold

Let M be a real, smooth, connected, finite-dimensional manifold. The pair (M, F) is called a *Finsler manifold* iff there exists a *fundamental function* $F : TM \rightarrow \mathbb{R}$, not necessary reversible (i.e., $F(x, -y)$ need not be equal to $F(x, y)$), that satisfies the following set of axioms (see, e.g., [Udriste and Neagu (1999)]):

- F1** $F(x, y) > 0$ for all $x \in M$, $y \neq 0$.
F2 $F(x, \lambda y) = |\lambda|F(x, y)$ for all $\lambda \in \mathbb{R}$, $(x, y) \in TM$.
F3 the *fundamental metric tensor* g_{ij} on M , given by

$$g_{ij}(x, y) = \frac{1}{2} \frac{\partial^2 F^2}{\partial y^i \partial y^j},$$

is *positive definite*.

- F4** F is *smooth* (C^∞) at every point $(x, y) \in TM$ with $y \neq 0$ and *continuous* (C^0) at every $(x, 0) \in TM$. Then, the *absolute Finsler energy function* is given by

$$F^2(x, y) = g_{ij}(x, y)y^i y^j.$$

Let $c = c(t) : [a, b] \rightarrow M$ be a smooth regular curve on M . For any two vector-fields $X(t) = X^i(t) \frac{\partial}{\partial x^i} \Big|_{c(t)}$ and $Y(t) = Y^i(t) \frac{\partial}{\partial x^i} \Big|_{c(t)}$ along the curve $c = c(t)$, we introduce the *scalar (inner) product* [Chern (1996)]

$$g(X, Y)(c) = g_{ij}(c, \dot{c})X^i Y^j$$

along the curve c .

In particular, if $X = Y$ then we have $\|X\| = \sqrt{g(X, X)}$. The vector-fields X and Y are *orthogonal* along the curve c , denoted by $X \perp Y$, iff $g(X, Y) = 0$.

Let $CT(N) = (L_{jk}^i, N_j^i, C_{jk}^i)$ be the *Cartan canonical N -linear metric connection* determined by the metric tensor $g_{ij}(x, y)$. The coefficients of

this connection are expressed by [Udriste and Neagu (1999)]

$$L_{jk}^i = \frac{1}{2}g^{im} \left(\frac{\delta g_{mk}}{\delta x^j} + \frac{\delta g_{jm}}{\delta x^k} - \frac{\delta g_{jk}}{\delta x^m} \right), \quad C_{jk}^i = \frac{1}{2}g^{im} \left(\frac{\partial g_{mk}}{\partial y^j} + \frac{\partial g_{jm}}{\partial y^k} - \frac{\partial g_{jk}}{\partial y^m} \right),$$

$$N_j^i = \frac{1}{2} \frac{\partial}{\partial y^j} (\Gamma_{kl}^i y^k y^l) = \frac{1}{2} \frac{\partial \Gamma_{00}^i}{\partial y^j}, \quad \Gamma_{jk}^i = \frac{1}{2}g^{im} \left(\frac{\partial g_{mk}}{\partial x^j} + \frac{\partial g_{jm}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^m} \right),$$

where $\frac{\delta}{\delta x^i} = \frac{\partial}{\partial x^i} + N_i^j \frac{\partial}{\partial y^j}$.

Let X be a vector-field along the curve c expressed locally by $X(t) = X^i(t) \frac{\partial}{\partial x^i} \Big|_{c(t)}$. Using the Cartan N -linear connection, we define the *co-variant derivative* $\frac{\nabla X}{dt}$ of $X(t)$ along the curve $c(t)$, by [Udriste and Neagu (1999)]

$$\frac{\nabla X}{dt} = \{ \dot{X}^i + X^m [L_{mk}^i(c, \dot{c}) \dot{c}^k + C_{mk}^i(c, \dot{c}) \frac{\delta}{\delta t}(\dot{c}^k)] \} \frac{\partial}{\partial x^i} \Big|_{c(t)}.$$

Since $\frac{\delta}{\delta t}(\dot{c}^k) = \ddot{c}^k + N_l^k(c, \dot{c}) \dot{c}^l$,

we have $\frac{\nabla X}{dt} = \{ \dot{X}^i + X^m [\Gamma_{mk}^i(c, \dot{c}) \dot{c}^k + C_{mk}^i(c, \dot{c}) \ddot{c}^k] \} \frac{\partial}{\partial x^i} \Big|_{c(t)}$ (3.157)

where $\Gamma_{mk}^i(c, \dot{c}) = L_{mk}^i(c, \dot{c}) + C_{ml}^i(c, \dot{c}) N_k^l(c, \dot{c})$.

In particular, c is a *geodesic* iff $\frac{\nabla \dot{c}}{dt} = 0$.

Since $CT(N)$ is a metric connection, we have

$$\frac{d}{dt} [g(X, Y)] = g \left(\frac{\nabla X}{dt}, Y \right) + g \left(X, \frac{\nabla Y}{dt} \right).$$

3.11.2 Energy Functional, Variations and Extrema

Let $x_0, x_1 \in M$ be two points not necessarily distinct. We introduce the Ω -set on M , as

$$\Omega = \{c : [0, 1] \rightarrow M \mid c \text{ is piecewise } C^\infty \text{ regular curve, } c(0) = x_0, c(1) = x_1\}.$$

For every $p \in \mathbb{R} - \{0\}$, we can define the p -energy functional on M

[Udriste and Neagu (1999)]

$$E_p : \Omega \rightarrow \mathbb{R}_+, \quad \text{as}$$

$$E_p(c) = \int_0^1 [g_{ij}(c, \dot{c}) \dot{c}^i \dot{c}^j]^{p/2} dt = \int_0^1 [g(\dot{c}, \dot{c})]^{p/2} dt = \int_0^1 \|\dot{c}\|^p dt.$$

In particular, for $p = 1$ we get the *length functional*

$$L(c) = \int_0^1 \|\dot{c}\| dt,$$

and for $p = 2$ we get the *energy functional*

$$E(c) = \int_0^1 \|\dot{c}\|^2 dt.$$

Also, for any naturally parametrized curve (i.e., $\|\dot{c}\| = \text{const}$) we have

$$E_p(c) = (L(c))^p = (E(c))^{p/2}.$$

Note that the p -energy of a curve is dependent of parametrization if $p \neq 1$.

For every curve $c \in \Omega$, we define the *tangent space* $T_c\Omega$ as

$T_c\Omega = \{X : [0, 1] \rightarrow TM \mid X \text{ is continuous, piecewise } C^\infty, X(t) \in T_{c(t)}M, \text{ for all } t \in [0, 1], X(0) = X(1) = 0\}.$

Let $(c_s)_{s \in (-\epsilon, \epsilon)} \subset \Omega$ be a *one-parameter variation* of the curve $c \in \Omega$. We define

$$X(t) = \frac{dc_s}{ds}(0, t) \in T_c\Omega.$$

Using the equality

$$g\left(\frac{\nabla \dot{c}_s}{\partial s}, \dot{c}_s\right) = g\left(\frac{\nabla}{\partial t}\left(\frac{\partial c_s}{\partial s}\right), \dot{c}_s\right),$$

we can prove the following Theorem: The first variation of the p -energy is

$$\begin{aligned} \frac{1}{p} \frac{dE_p(c_s)}{ds}(0) &= - \sum_t g(X, \Delta_t(\|\dot{c}\|^{p-2}\dot{c})) \\ &\quad - \int_0^1 \|\dot{c}\|^{p-4} g\left(X, \|\dot{c}\|^2 \frac{\nabla \dot{c}}{dt} + (p-2)g\left(\frac{\nabla \dot{c}}{dt}, \dot{c}\right) \dot{c}\right) dt, \end{aligned}$$

where $\Delta_t(\|\dot{c}\|^{p-2}\dot{c}) = (\|\dot{c}\|^{p-2}\dot{c})_{t+} - (\|\dot{c}\|^{p-2}\dot{c})_{t-}$ represents the jump of $\|\dot{c}\|^{p-2}\dot{c}$ at the discontinuity point $t \in (0, 1)$ [Udriste and Neagu (1999)].

The curve c is a *critical point* of E_p iff c is a geodesic.

In particular, for $p = 1$ the curve c is a *reparametrized geodesic*.

Now, let $c \in \Omega$ be a critical point for E_p (i.e., the curve c is a geodesic). Let $(c_{s_1 s_2})_{s_1, s_2 \in (-\epsilon, \epsilon)} \subset \Omega$ be a *two-parameter variation* of c . Using the notations:

$$X(t) = \frac{\partial c_{s_1 s_2}}{\partial s_1}(0, 0, t) \in T_c \Omega, \quad Y(t) = \frac{\partial c_{s_1 s_2}}{\partial s_2}(0, 0, t) \in T_c \Omega,$$

$$\|\dot{c}\| = v = \text{constant}, \quad \text{and} \quad I_p(X, Y) = \frac{\partial^2 E_p(c_{s_1 s_2})}{\partial s_1 \partial s_2}(0, 0),$$

we get the following Theorem: The second variation of the p -energy is [Udriste and Neagu (1999)]

$$\frac{1}{pv^{p-4}} I_p(X, Y) = - \sum_t g \left(Y, v^2 \Delta_t \left(\frac{\nabla X}{dt} \right) + (p-2)g \left(\Delta_t \left(\frac{\nabla X}{dt} \right), \dot{c} \right) \dot{c} \right) \\ - \int_0^1 g \left(Y, v^2 \left[\frac{\nabla}{dt} \frac{\nabla X}{dt} + R^2(X, \dot{c}) \dot{c} \right] + (p-2)g \left(\left[\frac{\nabla}{dt} \frac{\nabla X}{dt} + R^2(X, \dot{c}) \dot{c} \right], \dot{c} \right) \dot{c} \right) dt,$$

where $\Delta_t \left(\frac{\nabla X}{dt} \right) = \left(\frac{\nabla X}{dt} \right)_{t+} - \left(\frac{\nabla X}{dt} \right)_{t-}$ represents the jump of $\frac{\nabla X}{dt}$ at the discontinuity point $t \in (0, 1)$; also, if $R^l_{ijk}(c, \dot{c})$ represents the components of the *Finsler curvature tensor*, then

$$R^2(X, \dot{c}) \dot{c} = R^l_{ijk}(c, \dot{c}) \dot{c}^i \dot{c}^j X^k \frac{\partial}{\partial x^l} = R^l_{jk}(c, \dot{c}) \dot{c}^j X^k \frac{\partial}{\partial x^l}.$$

In particular, we have

$$R^i_{jk} = \frac{\delta N^i_j}{\delta x^k} - \frac{\delta N^i_k}{\delta x^j}, \quad \text{and} \quad R^i_{hjk} = \frac{\delta L^i_{hj}}{\delta x^k} - \frac{\delta L^i_{hk}}{\delta x^j} + L^s_{hj} L^i_{sk} - L^s_{hk} L^i_{sj} + C^i_{hs} R^s_{jk}.$$

Moreover, using the *Ricci identities* for the *deflection tensors*, we also have

$$R^i_{jk} = R^i_{mjk} y^m = R^i_{0jk}.$$

$I_p(X, Y) = 0$ (for all $Y \in T_c \Omega$) iff X is a *Jacobi field*, i.e.,

$$\frac{\nabla}{dt} \frac{\nabla X}{dt} + R^2(X, \dot{c}) \dot{c} = 0.$$

In these conditions we have the following definition: A point $c(b)$ ($0 \leq a < b < 1$) of a geodesic $c \in \Omega$ is called a *conjugate point* of a point $c(a)$ along the curve $c(t)$, if there exists a non-zero Jacobi field which vanishes at $t \in \{a, b\}$.

Now, integrating by parts and using the property of metric connection, we find

$$\begin{aligned} \frac{1}{pv^{p-4}}I_p(X, Y) &= \int_0^1 v^2 \left[g \left(\frac{\nabla X}{dt}, \frac{\nabla Y}{dt} \right) - R^2(X, \dot{c}, Y, \dot{c}) \right] \\ &\quad + (p-2)g \left(\dot{c}, \frac{\nabla X}{dt} \right) g \left(\dot{c}, \frac{\nabla Y}{dt} \right) dt, \end{aligned}$$

where $R^2(X, \dot{c}, Y, \dot{c}) = g(R^2(Y, \dot{c})\dot{c}, X) = R_{0i0j}(c(t), \dot{c}(t))X^iY^j$.

Let $R_{ijk} = g_{jm}R_{ik}^m$. In any Finsler space the following identity is satisfied,

$$R_{ijk} + R_{jki} + R_{kij} = 0,$$

get by the *Bianchi identities*. As $R_{0i0j} = R_{i0j} = R_{j0i} = R_{0j0i}$ we get $R^2(X, \dot{c}, Y, \dot{c}) = R^2(Y, \dot{c}, X, \dot{c})$.

The quadratic form associated to the *Hessian of the p -energy* is given by

$$I_p(X) = I_p(X, X) = \int_0^1 v^2 \left[\left\| \frac{\nabla X}{dt} \right\|^2 - R^2(X, \dot{c}, X, \dot{c}) \right] + (p-2) \left[g \left(\dot{c}, \frac{\nabla X}{dt} \right) \right]^2 dt.$$

Let

$$T_c^\perp \Omega = \{X \in T_c \Omega \mid g(X, \dot{c}) = 0\}, \quad \text{and}$$

$T_c' \Omega = \{X \in T_c \Omega \mid X = f\dot{c}, \text{ where } f : [0, 1] \rightarrow \mathbb{R} \text{ is continuous, piecewise } C^\infty, f(0) = f(1) = 0\}.$

Let c be a geodesic and $p \in \mathbb{R} - \{0, 1\}$. Then $I_p(T_c' \Omega) \geq 0$ for $p \in (-\infty, 0) \cup (1, \infty)$, and $I_p(T_c' \Omega) \leq 0$ for $p \in (0, 1)$. Moreover, in both cases: $I_p(X) = 0$ iff $X = 0$. To prove it, let $X = f\dot{c} \in T_c' \Omega$. Then we have [Udriste and Neagu (1999)]

$$\begin{aligned} \frac{1}{v^{p-4}}I_p(X) &= p \int_0^1 \left\{ v^2 [g(f'\dot{c}, f'\dot{c}) - R^2(f\dot{c}, \dot{c}, f\dot{c}, \dot{c})] + (p-2) [g(\dot{c}, f'\dot{c})]^2 \right\} dt \\ &= p \int_0^1 [v^4(f')^2 + (p-2)v^4(f')^2] dt = \int_0^1 p(p-1)v^4(f')^2 dt. \end{aligned}$$

Moreover, if $I_p(X) = 0$, then $f' = 0$, which means that f is constant. The conditions $f(0) = f(1) = 0$ imply that $f = 0$.

As $I_p(T'_c\Omega)$ is positive definite for $p \in (-\infty, 0) \cup (1, \infty)$ and negative definite for $p \in (0, 1)$, it is sufficient to study the behavior of I_p restricted to $T_c^\perp\Omega$. Since $X \perp \dot{c}$ and the curve c is a geodesic it follows

$$g\left(\dot{c}, \frac{\nabla X}{dt}\right) = 0.$$

Hence, for all $X \in T_c^\perp\Omega$, we have

$$\frac{1}{pv^{p-2}}I_p(X) = \int_0^1 \left[\left\| \frac{\nabla X}{dt} \right\|^2 - R^2(X, \dot{c}, X, \dot{c}) \right] dt = I(X).$$

3.11.3 Application: *Finsler–Lagrangian Field Theory*

In this subsection we present *generalized Finsler–Lagrangian field theory*. The geometrical background of this theory relies on the notion of *generalized Lagrangian space*, $GL^n = (M, g_{ij}(x^k, y^k))$, which is a real n D manifold M with local coordinates $\{x^i\}$, ($i = 1, \dots, n$) and a symmetric *fundamental metric tensor-field* $g_{ij} = g_{ij}(x^k, y^k)$ of rank n and constant signature on T [Miron *et. al.* (1988); Miron and Anastasiei (1994)].

From physical point of view, the fundamental metric tensor represents a *unified* gravitational field on TM , which consists of one *external* (x)–gravitational field spanned by points $\{x^i\}$, and the one *internal* (y)–gravitational field spanned by directions $\{y^i\}$ and equipped with some microscopic character of the space–time structure.

The field theory developed on a generalized Lagrangian space GL^n relies on a fixed *a priori* nonlinear connection $\Gamma = (N_j^i(x, y))$ on the tangent bundle TM . This plays the role of mapping operator of the internal (y)–field onto the external (x)–field, and prescribes the *interaction* between (x)– and (y)–fields. From geometrical point of view, the nonlinear connection allows the construction of the *adapted bases* [Miron *et. al.* (1988); Miron and Anastasiei (1994)]

$$\left\{ \frac{\delta}{\delta x^i} = \frac{\partial}{\partial x^i} - N_j^i \frac{\partial}{\partial y^j}, \frac{\partial}{\partial y^i} \right\} \subset \mathcal{X}(TM),$$

$$\{dx^i, \delta y^i = dy^i + N_j^i dx^j\} \subset \mathcal{X}^*(TM).$$

As to the spatial structure, the most important thing is to determine the *Cartan canonical connection* $CT = (L_{jk}^i, C_{jk}^i)$ with respect to g_{ij} , which

comes from the metric conditions

$$g_{ij|k} = \frac{\delta g_{ij}}{\delta x^k} - L_{ik}^m g_{mj} - L_{jk}^m g_{mi} = 0, \quad g_{ij;k} = \frac{\partial g_{ij}}{\partial y^k} - C_{ik}^m g_{mj} - C_{jk}^m g_{mi} = 0,$$

where “ $|_k$ ” and “ $_{;k}$ ” are the local h - and v - covariant derivatives of CT . The importance of the Cartan connection comes from its main role played in the generalized *Finsler-Lagrangian theory* of physical fields.

Regarding the *unified field* $g_{ij}(x, y)$ of GL^n , the authors of [Miron *et. al.* (1988); Miron and Anastasiei (1994)] constructed a *Sasakian metric* on TM ,

$$G = g_{ij} dx^i \otimes dx^j + g_{ij} \delta y^i \otimes \delta y^j.$$

In this context, the *Einstein equations* for the gravitational potentials $g_{ij}(x, y)$ of a generalized Lagrangian space GL^n , ($n > 2$), are postulated as being the Einstein equations attached to CT and G ,

$$\begin{aligned} R_{ij} - \frac{1}{2} R g_{ij} &= K T_{ij}^H, & 'P_{ij} &= K T_{ij}^1, \\ S_{ij} - \frac{1}{2} S g_{ij} &= K T_{ij}^V, & ''P_{ij} &= -K T_{ij}^2, \end{aligned}$$

where $R_{ij} = R_{ijm}^m$, $S_{ij} = S_{ijm}^m$, $'P_{ij} = P_{ijm}^m$, $''P_{ij} = P_{imj}^m$ are the Ricci tensors of CT , $R = g^{ij} R_{ij}$ and $S = g^{ij} S_{ij}$ are the scalar curvatures, T_{ij}^H , T_{ij}^V , T_{ij}^1 , T_{ij}^2 are the components of the energy-momentum tensor T , and K is the Einstein constant (equal to 0 for vacuum). Moreover, the energy-momentum tensors T_{ij}^H and T_{ij}^V satisfy the *conservation laws* [Miron *et. al.* (1988); Miron and Anastasiei (1994)]

$$K T_{j|m}^H{}^m = -\frac{1}{2} (P_{js}^{hm} R_{hm}^s + 2 R_{mj}^s P_s^m), \quad K T_{j|m}^V{}^m = 0.$$

The generalized Lagrangian theory of electromagnetism relies on the *canonical Liouville vector-field* $C = y^i \frac{\partial}{\partial y^i}$ and the Cartan connection CT of the generalized Lagrangian space GL^n . In this context, we can introduce the *electromagnetic two-form* on TM [Miron and Anastasiei (1994)]

$$\begin{aligned} F &= F_{ij} \delta y^i \wedge dx^j + f_{ij} \delta y^i \wedge \delta y^j, & \text{where} \\ F_{ij} &= \frac{1}{2} [(g_{im} y^m)_{|j} - (g_{jm} y^m)_{|i}], & f_{ij} = \frac{1}{2} [(g_{im} y^m)_{;j} - (g_{jm} y^m)_{;i}]. \end{aligned}$$

Using the Bianchi identities attached to the Cartan connection CT , they conclude that the electromagnetic components F_{ij} and f_{ij} are governed by

the following *Maxwell-type equations*

$$\begin{aligned} F_{ij|k} + F_{jk|i} + F_{ki|j} &= -[C_{imr}y^m + (g_{im}y^m)_{|r}]R_{jk}^r, \\ F_{ij;k} + F_{jk;i} + F_{ki;j} &= -(f_{ij|k} + f_{jk|i} + f_{ki|j}), \quad f_{ij;k} + f_{jk;i} + f_{ki;j} = 0. \end{aligned}$$

3.11.4 Riemann–Finsler Approach to Information Geometry

3.11.4.1 Model Specification and Parameter Estimation

Model as a Parametric Family of Probability Distributions

From a statistical standpoint, observed data is a random sample from an unknown population. Ideally, the goal of modelling is to deduce the population that generated the observed data. Formally, a model is defined as a parametric family of probability distributions (see [Myung and Pitt (2003)]).

Let us use $f(y|w)$ to denote the *probability distribution function* that gives the probability of observing *data*, $y = (y_1, \dots, y_m)$, given the *model's parameter vector*, $w = (w_1, \dots, w_k)$. Under the assumption that individual observations, y_i 's, are independent of one another, $f(y|w)$ can be rewritten as a product of individual probability distribution functions,

$$f(y = (y_1, \dots, y_m) | w) = f(y_1 | w) f(y_2 | w) \dots f(y_m | w). \quad (3.158)$$

Parameter Estimation

Once a model is specified with its parameters and data have been collected, the model's ability to fit the data can be assessed. Model fit is measured by finding parameter values of the model that give the 'best' fit to the data in some defined sense – a procedure called *parameter estimation* in statistics.

There are two generally accepted methods of parameter estimation: *least-squares estimation* (LSE) and *maximum likelihood estimation* (MLE). In LSE, we seek the parameter values that *minimize the sum of squares error* (SSE) between observed data and a model's predictions:

$$SSE(w) = \sum_{i=1}^m (y_i - y_{i,prd}(w))^2,$$

where $y_{i,prd}(w)$ denotes the model's prediction for observation y_i . In MLE, we seek the parameter values that are most likely to have produced the data.

This is obtained by *maximizing* the *log-likelihood* of the observed data:

$$\mathbf{loglik}(w) = \sum_{i=1}^m \ln f(y_i|w).$$

By maximizing either the likelihood or the log-likelihood, the same solution is obtained because the two are monotonically related to each other. In practice, the log-likelihood is preferred for computational ease. The parameters that minimize the sum of squares error or the log-likelihood are called the LSE or MLE *estimates*, respectively.

For normally distributed data with constant variance, LSE and MLE are equivalent in the sense that both methods yield the same parameter estimates. For non-normal data such as proportions and response times, however, LSE estimates tend to differ from MLE estimates. Although LSE is often the ‘de facto’ method of estimation in cognitive psychology, MLE is a preferred method of estimation in statistics, especially for non-normal data. In particular, MLE is well-suited for statistical inference in hypothesis testing and model selection. Finding LSE or MLE estimates generally requires use of a numerical optimization procedure.

3.11.4.2 Model Evaluation and Testing

Qualitative Criteria

A model satisfies the *explanatory adequacy criterion* if its assumptions are plausible and consistent with established findings, and importantly, the theoretical account is reasonable for the cognitive process of interest. In other words, the model must be able to do more than redescribe observed data. The model must also be interpretable in the sense that the model makes sense and is understandable. Importantly, the components of the model, especially, its parameters, must be linked to psychological processes and constructs. Finally, the model is said to be faithful to the extent that the model’s ability to capture the underlying mental process originates from the theoretical principles embodied in the model, rather than from the choices made in its computational instantiation.

3.11.4.3 Quantitative Criteria

Falsifiability

This is a *necessary condition* for testing a model or theory, refers to *whether there exist potential observations that a model cannot describe* [Pop-

per (1959)]. If so, then the model is said to be falsifiable. An unfalsifiable model is one that can describe all possible data patterns in a given experimental situation. *There is no point in testing an unfalsifiable model.*

A heuristic rule for determining a model's falsifiability is already familiar to us: The model is falsifiable if the number of its free parameters is less than the number of data observations. This *counting rule*, however, turns out to be imperfect, in particular, for nonlinear models. To remedy limitations of the counting rule, [Bamber and Santen (1985)] provided a *formal rule* for assessing a model's falsifiability, which yielded the counting rule as a special case. The rule states that a model is falsifiable if the *rank of its Jacobian matrix* is less than the number of data observations for all values of the parameters. Recall that the Jacobian matrix is defined in terms of partial derivatives as: $J_{ij}(w) = \partial E(y_j) / \partial w_i$ ($i = 1, \dots, k; j = 1, \dots, m$) where $E(x)$ stands for the *expectation* of a random variable x .

Goodness of Fit

A model should also give a good description of the observed data. Goodness of fit refers to the *model's ability to fit* the particular set of observed data. Common examples of *goodness of fit measures* are the minimized sum of squares error (SSE), the mean squared error (MSE), the root mean squared error (RMSE), the percent variance accounted for (PVAF), and the maximum likelihood (ML).

The first four of these, defined below, are related to one another in a way that one can be written in terms of another:

$$\begin{aligned} MSE &= SSE(w_{LSE}^*) / m, \\ RMSE &= \sqrt{SSE(w_{LSE}^*) / m}, \\ PVAF &= 100(1 - SSE(w_{LSE}^*) / SST), \\ ML &= f(y|w_{MLE}^*). \end{aligned}$$

Here w_{LSE}^* is the parameter that minimizes $SSE(w)$, that is, an LSE estimate, and SST stands for the sum of squares total defined as $SST = \sum_i (y_i - y_{mean})^2$. ML is the probability distribution function maximized with respect to the model's parameters, evaluated at w_{MLE}^* , which is obtained through MLE.

Complexity

Not only should a model describe the data in hand well, but it should also do so in the least complex (i.e., simplest) way. Intuitively, complexity has to do with a model's inherent flexibility that enables it to fit a wide range of data patterns. There seem to be at least *two dimensions of model complexity*, the *number of parameters* and the model's *functional form*. The latter refers to the way the parameters are combined in the model equation. The more parameters a model has, the more complex it is. Importantly also, two models with the same number of parameters but different functional forms can differ significantly in their complexity. For example, it seems unlikely that two one-parameter models, $y = x + w$ and $y = e^{wx}$ are equally complex. The latter is probably much better at fitting data than the former.

It turns out that one can devise a quantitative measure of model complexity that takes into account both dimensions of complexity and at the same time is theoretically justified as well as intuitive. One example is the *geometric complexity* (GC) of a model [Pitt *et. al.* (2002)] defined as:

$$GC = \frac{k}{2} \ln \frac{n}{2\pi} + \ln \int dw \sqrt{\det I(w)}, \quad (3.159)$$

or

$$GC = \text{parametric complexity} + \text{functional complexity},$$

where k is the number of parameters, n is the sample size, $I(w)$ is the *Fisher information matrix* (or, *covariance matrix*) defined as

$$I_{ij}(w) = -E \left[\partial^2 \ln f(y|w) / \partial w_i \partial w_j \right], i, j = 1, \dots, k, \quad (3.160)$$

or

$$I_{ij} = -\text{Expect.Value}(\text{Hessian}(\mathbf{loglik}(w))).$$

Functional form effects of complexity are reflected in the second term of GC through $I(w)$. How do we interpret geometric complexity? The meaning of geometric complexity is related to the *number of 'different'* (i.e., distinguishable) *probability distributions* that a model can account for. The more distinguishable distributions that the model can describe by finely tuning its parameter values, the more complex it is ([Myung *et. al.* (2000a)]). For example, when geometric complexity is calculated for the following two-parameter psychophysical models, *Stevens' law* ($y = w_1 x^{w_2}$) and *Fechner's logarithmic law* ($y = w_1 \ln(x + w_2)$), the former turns out to be more complex than the latter [Pitt *et. al.* (2002)].

Generalizability

The fourth quantitative criterion for model evaluation is generalizability. This criterion is defined as a *model's ability to fit not only the observed data at hand, but also new, as yet unseen data samples from the same probability distribution*. In other words, model evaluation should not be focused solely on how well a model fits observed data, but how well it fits data generated by the cognitive process underlying the data. This goal will be achieved best when generalizability is considered.

To summarize, these four quantitative criteria work together to assist in model evaluation and guide (even constrain) model development and selection. The model must be sufficiently complex, but not too complex, to capture the regularity in the data. Both a good fit to the data and good generalizability will ensure an appropriate degree of complexity, so that the model captures the regularity in the data. In addition, because of its broad focus, generalizability will constrain the power of the model, thus making it falsifiable. Although all four criteria are inter-related, generalizability may be the most important. By making it the guiding principle in model evaluation and selection, one cannot go wrong.

3.11.4.4 Selection Among Different Models

Since a model's generalizability is not directly observable, it must be estimated using observed data. The measure developed for this purpose trades off a model's fit to the data with its complexity, the aim being to select the model that is complex enough to capture the regularity in the data, but not overly complex to capture the ever-present random variation. Looked at in this way, generalizability embodies the principle of *Occam's razor* (or *principle of parsimony*, i.e., the requirement of maximal simplicity of cognitive models).

Model Selection Methods

Now, we describe specific measures of generalizability. Four representative *generalizability criteria* are introduced. They are the *Akaike Information Criterion* (AIC), the *Bayesian Information Criterion* (BIC), *crossvalidation* (CV), and *minimum description length* (MDL) (see a special Journal of Mathematical Psychology issue on *model selection*, in particular [Myung et. al. (2000b)]). In all four methods, the *maximized log-likelihood* is used as a *goodness-of-fit measure*, but they differ in how model complexity is conceptualized and measured.

AIC and BIC

AIC and BIC for a given model are defined as follows:

$$\begin{aligned}AIC &= -2 \ln f(y|w^*) + 2k, \\BIC &= -2 \ln f(y|w^*) + k \ln n,\end{aligned}$$

where w^* is a MLE estimate, k is the number of parameters and n is the sample size. For normally distributed errors with constant variance, the first term of both criteria, $-2 \ln f(y|w^*)$, is reduced to $(n \cdot \ln(SSE(w^*)) + c_o)$ where c_o is a constant that does not depend upon the model. In each criterion, the first term represents a lack of fit measure, the second term represents a complexity measure, and together they represent a lack of generalizability measure. A lower value of the criterion means better generalizability. Therefore, the model that minimizes a given criterion should be chosen.

Complexity in AIC and BIC is a *function of only the number of parameters*. Functional form, another important dimension of model complexity, is not considered. For this reason, these methods are not recommended for comparing models with the same number of parameters but different functional forms. The other two selection methods, CV and MDL, described next, are sensitive to functional form as well as the number of parameters.

Cross-Validation

In CV, a model's generalizability is estimated without defining an explicit measure of complexity. Instead, models with more complexity than necessary to capture the regularity in the data are penalized through a *resampling procedure*, which is performed as follows: The observed data sample is divided into two sub-samples, *calibration* and *validation*. The calibration sample is then used to find the best-fitting values of a model's parameters by MLE or LSE. These values, denoted by w_{cal}^* , are then fixed and fitted, without any further tuning of the parameters, to the validation sample, denoted by y_{val} . The resulting fit to y_{val} by w_{cal}^* is called as the model's CV index and is taken as the model's generalizability estimate. If desired, this single-division-based CV index may be replaced by the average CV index calculated from multiple divisions of calibration and validation samples. The latter is a more accurate estimate of the model's generalizability, though it is also more computationally demanding.

The main attraction of cross-validation is its *ease of use*. All that is needed is a simple resampling routine that can easily be programmed on any desktop computer. The second attraction is that unlike AIC and BIC, CV is sensitive to the functional form dimension of model complexity, though how it works is unclear because of the implicit nature of the method. For these reasons, the method can be used in all modelling situations, including the case of comparing among models that differ in functional form but have the same number of parameters.

Minimum Description Length

MDL is a selection method that has its origin in *algorithmic coding theory* in computer science. According to MDL, both models and data are viewed as codes that can be compressed. The basic idea of this approach is that regularities in data necessarily imply the existence of *statistical redundancy* and that the redundancy can be used to compress the data [Grunwald (1999); Grunwald (2000); Grunwald *et. al.* (2005)]. Put another way, the amount of *regularity in data* is directly related to the *data description length*. The shorter the description of the data by the model, the better the approximation of the underlying regularity, and thus, the higher the model's generalizability is. Formally, MDL is defined as:

$$MDL = -\ln f(y|w^*) + \frac{k}{2} \ln \frac{n}{2\pi} + \ln \int dw \sqrt{\det I(w)}, \quad (3.161)$$

or

$$MDL = \text{lack-of-fit measure} + \text{param. complexity} + \text{functional complexity},$$

where the first term is the same lack of fit measure as in AIC and BIC; the second and third terms together represent the geometric complexity measure (3.159). In coding theory, MDL is interpreted as the length in bits of the shortest possible code that describes the data unambiguously with the help of a model. The *model with the minimum value of MDL* encodes the most regularity in the data, and therefore should be preferred.

The second term in (3.161), which captures the effects of model complexity due to the number of parameter (k), is a logarithmic function of sample size n . In contrast, the third term, which captures functional form effects, is not sensitive to sample size. This means that as sample size increases, the relative contribution of the effects due to functional form to those due to the number of parameters will be gradually reduced. Therefore, *functional form effects can be ignored for sufficiently large n* , in which

case the MDL value becomes approximately equal to one half of the BIC value.

Probably the most desirable property of MDL over other selection methods is that its complexity measure takes into account the effects of both dimensions of model complexity, the number of parameters and functional form. The MDL complexity measure, unlike CV, shows explicitly how both factors contribute to model complexity. In short, *MDL is a sharper and more accurate method* than these three competitors. The price that is paid for MDL's superior performance is its *computational cost*. MDL can be laborious to calculate. First, the Fisher information matrix (3.160) must be obtained by calculating the second derivatives (i.e., Hessian matrix) of the log-likelihood function, $\ln f(y|w)$. This calculation can be non-trivial, though not impossible. Second, the square-root of the determinant of the Fisher information matrix must be integrated over the parameter space. This generally requires use of a numerical integration method such as *Markov Chain Monte Carlo* (see e.g., [Gilks *et. al.* (1996)]).

3.11.4.5 *Riemannian Geometry of Minimum Description Length*

From a geometric perspective, a *parametric model family of probability distributions* (3.158) forms a *Riemannian manifold embedded in the space of all probability distributions* (see [Amari (1985); Amari and Nagaoka (2000); McCullagh (1987)]). *Every distribution is a point in this space*, and the *collection of points created by varying the parameters of the model induces a manifold* in which 'similar' distributions are mapped to 'nearby' points. The *infinitesimal distance between points* separated by the infinitesimal parameter differences dw^i is given by

$$ds^2 = g_{ij}(w) dw^i dw^j,$$

where $g_{ij}(w)$ is the *Riemannian metric tensor*. The Fisher information, $I_{ij}(w)$, defined by (3.160), is the natural metric on a manifold of distributions in the context of statistical inference [Amari (1985)]. We argue that the MDL measure of model fitness has an attractive interpretation in such a geometric context.

The first term in MDL equation (3.161) estimates the accuracy of the model since the likelihood $f(y|w^*)$ measures the ability of the model to fit the observed data. The second and third terms are supposed to penalize model complexity; we will show that they have interesting geometric inter-

pretations. Given the metric $I_{ij}(w) = g_{ij}(w)$ on the space of parameters, the *infinitesimal volume element* on the parameter manifold is

$$dV = dw \sqrt{\det I(w)} = \prod_{l=1}^k dw^l \sqrt{\det I(w)}.$$

The *Riemannian volume of the parameter manifold* is obtained by integrating dV over the space of parameters:

$$V_M = \int dV = \int dw \sqrt{\det I(w)}.$$

In other words, the *third term (functional complexity) in MDL penalizes models that occupy a large volume in the space of distributions*.

In fact, the volume measure V_M is related to the number of ‘distinguishable’ probability distributions indexed by the model. Because of the way the model family is embedded in the space of distributions, two different parameter values can index very similar distributions. If complexity is related to volumes occupied by model manifolds, the measure of volume should count only different, or distinguishable, distributions, and not the artificial coordinate volume. It is shown in [Myung *et al.* (2000a)] that the volume V_M achieves this goal.

Selecting Among Qualitative Models

Application of any of the preceding selection methods requires that the models are *quantitative models*, each defined as a *parametric family of probability distributions*.

Pseudo-probabilistic MDL Approach

The ‘pseudo-probabilistic’ approach [Grunwald (1999)] for selecting among qualitative models derives a selection criterion that is similar to the MDL criterion for quantitative models, but it is a formulation that is closer to the original spirit of the MDL principle, which states:

‘Given a data set D and a model M , the description length of the data, $DL_M(D)$, is given by the sum of (a) the description length of the data when encoded with help of the model, $DL(D|M)$, and (b) the description length of the model itself, $DL(M)$: $DL_M(D) = DL(D|M) + DL(M)$. Among a set of competing models, the best model is the one that minimizes $DL_M(D)$.’

The above MDL principle is broad enough to include the MDL criterion for quantitative models as a specific instantiation. The first, lack-of-fit

term of the quantitative criterion $(-\ln f(y|w^*))$ can be seen as $DL(D|M)$, whereas the second and third terms $(\frac{k}{2} \ln \frac{n}{2\pi} + \ln \int dw \sqrt{\det I(w)})$ represent geometric complexity as $DL(M)$. Likewise, a computable criterion that implements the above principle can be obtained with the pseudo-probabilistic approach. It is derived from the *Kraft–Inequality Theorem* in coding theory ([Li and Vitanyi (1997)]). The Theorem proves that one can always associate arbitrary models with their ‘equivalent’ probability distributions in a procedure called *entropification* [Grunwald (1999)].

MDL Criterion for Qualitative Models

Entropification proceeds as follows. We first ‘construct’ a parametric family of probability distributions for a given qualitative model in the following form:

$$p(y = (y_1, \dots, y_m) | w) = \exp \left(-w \sum_{i=1}^m \text{Err}(y_{i, \text{obs}} - y_{i, \text{prd}}(w)) \right) / Z(w).$$

In this equation, $\text{Err}(x)$ is an error function that measures the model’s prediction performance such as $\text{Err}(x) = |x|$ or x^2 , w is a scalar parameter, and $Z(w)$ is the normalizing factor defined as

$$Z(w) = \sum_{y_1} \dots \sum_{y_m} p(y = (y_1, \dots, y_m) | w).$$

The above formulation requires that each observation y_i be represented by a discrete variable that takes on a finite number of possible values representing the model’s qualitative (e.g., ordinal) predictions.

Once a suitable error function, $\text{Err}(x)$, is chosen, the above probability distribution function is then used to fit observed data, and the best-fitting parameter w^* is sought by MLE. The description length of the data encoded with the help of the model is then obtained by taking the minus logarithm of the maximum likelihood (ML),

$$DL(D|M) = -\ln p(y|w^*).$$

The second term, $DL(M)$, the description length of the model itself, is obtained by counting the number of different data patterns the model can account for and then taking the logarithm of the resulting number. Putting these together, the desired MDL criterion for a qualitative model is given

by

$$MDL_{qual} = w^* \sum_{i=1}^m \text{Err}(y_{i,obs} - y_{i,prd}(w^*)) + \ln Z(w^*) + \ln N,$$

where N is the number of all possible data patterns or data sets that the model predicts.

3.11.4.6 Finsler Approach to Information Geometry

Recall that information geometry has emerged from investigating the geometrical structure of a *family of probability distributions*, and has been applied successfully to various areas including statistical inference, control theory and multi-terminal information theory (see [Amari (1985); Amari and Nagaoka (2000)]). In this subsection we give a brief review on a more general approach to information geometry, based on Finsler geometry (see subsection 3.11 above).

A *parameter-space of probability distributions*, defined by

$$M = \{x : p = p(r, x) \text{ is a probability distribution on } \mathbb{R}\}$$

represents a smooth manifold, called the *probability manifold* [Amari (1985); Amari and Nagaoka (2000)]. On a probability manifold M we can define a *probability divergence* $D = D(x, y)$, as

$$D(x, y) = \int_M p(r, x) f\left(\frac{p(r, y)}{p(r, x)}\right) dr,$$

where $f(\cdot)$ is a convex function such that $f(1) = 0$, $f''(1) = 1$,

which satisfies the following conditions [Shen (2005)]

$$\begin{aligned} D(x, y) &> 0 & \text{if } & x \neq y, \\ D(x, y) &= 0 & \text{if } & x = y, \\ D(x, y) &\neq D(y, x) & \text{in general.} \end{aligned}$$

On the other hand, if $d = d(x, y)$ is a *probability distance* on a probability manifold M , satisfying the following standard conditions:

$$\begin{aligned} d(x, x) &= 0, \\ d(x, y) &> 0, & \text{if } & x \neq y, \\ d(x, y) &\leq d(x, r) + d(r, y) & \text{(triangle inequality),} \end{aligned}$$

then for any function $\psi = \psi(h)$ with $\psi(0) = 0, \psi(h) > 0$ for $h > 0$, the probability divergence on M is defined as

$$D(x, y) := \psi(d(x, y)). \quad (3.162)$$

Recall that a *Finsler metric* $L = L(x, y)$ is a function of tangent vectors y at a point $x \in M$, with the following properties:

$$\begin{aligned} L(x, ty) &= t^2 L(x, y) \quad \text{for } t > 0, \\ g_{ij}(x, y) &:= \frac{1}{2} \frac{\partial^2 L}{\partial y^i \partial y^j}(x, y) > 0, \\ F_x(y) &:= \sqrt{L(x, y)}, \quad F_x(u + v) \leq F_x(u) + F_x(v). \end{aligned} \quad (3.163)$$

This means that there is an inner product g_y at a point $x \in M$, such that

$$g_y(u, v) = g_{ij}(x, y) u^i v^j,$$

so that our Finsler metric $L(x, y) \in M$, given by (3.163), becomes

$$L(x, y) = g_y(u, v) = g_{ij}(x) y^i y^j.$$

Therefore, in a special case when $g_{ij}(x, y) = g_{ij}(x)$ are independent of y , the Finsler metric $L(x, y)$ becomes a standard Riemannian metric $g_{ij}(x) y^i y^j$. In this way, all the material from the previous subsection can be generalized to Finsler geometry.

Now, $D(x, y) \in M$, given by (3.162), is called the *regular divergence*, if

$$2D(x, x + y) = L(x, y) + \frac{1}{2} L_{x^k}(x, y) y^k + \frac{1}{3} H(x, y) + o(|y|^3),$$

where $H = H(x, y) \in M$ is homogenous function of degree 3 in y , i.e.,

$$H(x, ty) = t^3 H(x, y) \quad \text{for } t > 0.$$

A pair $\{L, H\} \in M$ is called a *Finsler information structure* [Shen (2005)].

In a particular case when $L(x, y) = g_{ij}(x) y^i y^j$ is a Riemannian metric, and $H(x, y) = H_{ijk}(x) y^i y^j y^k$ is a polynomial, then we have *affine information structure* $\{L, H\} \in M$, which is described by a family of affine connections, called α -connections by [Amari (1985); Amari and Nagaoka (2000)]. However, in general, the induced information structure $\{L, H\} \in M$ is not affine, i.e., $L(x, y)$ is not Riemannian and $H(x, y)$ is not polynomial.

3.12 Symplectic Manifolds and Their Applications

3.12.1 Symplectic Algebra

Symplectic algebra works in the category of symplectic vector spaces V_i and linear symplectic mappings $t \in L(V_i, V_j)$ [Puta (1993)].

Let V be a n D real vector space and $L^2(V, \mathbb{R})$ the space of all bilinear maps from $V \times V$ to \mathbb{R} . We say that a *bilinear map* $\omega \in L^2(V, \mathbb{R})$ is *nondegenerate*, i.e., if $\omega(v_1, v_2) = 0$ for all $v_2 \in V$ implies $v_1 = 0$.

If $\{e_1, \dots, e_n\}$ is a basis of V and $\{e^1, \dots, e^n\}$ is the dual basis, $\omega_{ij} = \omega(e_i, e_j)$ is the matrix of ω . A bilinear map $\omega \in L^2(V, \mathbb{R})$ is nondegenerate iff its matrix ω_{ij} is nonsingular. The transpose ω^t of ω is defined by $\omega^t(e_i, e_j) = \omega(e_j, e_i)$. ω is *symmetric* if $\omega^t = \omega$, and *skew-symmetric* if $\omega^t = -\omega$.

Let $A^2(V)$ denote the space of skew-symmetric bilinear maps on V . An element $\omega \in A^2(V)$ is called a 2-form on V . If $\omega \in A^2(V)$ is nondegenerate then in the basis $\{e_1, \dots, e_n\}$ its matrix $\omega(e_i, e_j)$ has the form

$$J = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}.$$

A *symplectic form* on a real vector space V of dimension $2n$ is a nondegenerate 2-form $\omega \in A^2(V)$. The pair (V, ω) is called a symplectic vector space. If (V_1, ω_1) and (V_2, ω_2) are symplectic vector spaces, a linear map $t \in L(V_1, V_2)$ is a *symplectomorphism* (i.e., a symplectic mapping) iff $t^*\omega_2 = \omega_1$. If (V, ω) is a symplectic vector space, we have an orientation Ω_ω on V given by

$$\Omega_\omega = \frac{(-1)^{\frac{n(n-1)}{2}}}{n!} \omega^n.$$

Let (V, ω) be a $2n$ D symplectic vector space and $t \in L(V, V)$ a symplectomorphism. Then t is volume preserving, i.e., $t^*(\Omega_\omega) = \Omega_\omega$, and $\det_{\Omega_\omega}(t) = 1$.

The set of all symplectomorphisms $t : V \rightarrow V$ of a $2n$ D symplectic vector space (V, ω) forms a group under composition, called the *symplectic group*, denoted by $Sp(V, \omega)$.

In matrix notation, there is a basis of V in which the matrix of ω is $J = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}$, such that $J^{-1} = J^t = -J$, and $J^2 = -I$. For $t \in L(V, V)$ with matrix $T = [T_j^i]$ relative to this basis, the condition $t \in Sp(V, \omega)$, i.e.,

$t^*\omega = \omega$, becomes

$$T^t J T = J.$$

In general, by definition a matrix $A \in M_{2n \times 2n}(\mathbb{R})$ is symplectic iff $A^t J A = J$.

Let (V, ω) be a symplectic vector space, $t \in Sp(V, \omega)$ and $\lambda \in \mathbb{C}$ an eigenvalue of t . Then λ^{-1} , $\bar{\lambda}$ and $\bar{\lambda}^{-1}$ are eigenvalues of t .

3.12.2 Symplectic Geometry

Symplectic geometry is a globalization of symplectic algebra [Putz (1993)]; it works in the category **Symplec** of symplectic manifolds M and symplectic diffeomorphisms f . The phase-space of a conservative dynamical system is a symplectic manifold, and its time evolution is a one-parameter family of symplectic diffeomorphisms.

A *symplectic form* or a symplectic structure on a smooth (i.e., C^k) manifold M is a nondegenerate closed 2-form ω on M , i.e., for each $x \in M$ $\omega(x)$ is nondegenerate, and $d\omega = 0$. A *symplectic manifold* is a pair (M, ω) where M is a smooth $2n$ D manifold and ω is a symplectic form on it. If (M_1, ω_1) and (M_2, ω_2) are symplectic manifolds then a smooth map $f : M_1 \rightarrow M_2$ is called *symplectic map* or *canonical transformation* if $f^*\omega_2 = \omega_1$.

For example, any symplectic vector space (V, ω) is also a symplectic manifold; the requirement $d\omega = 0$ is automatically satisfied since ω is a constant map. Also, any orientable, compact surface Σ is a symplectic manifold; any non-vanishing 2-form (volume element) ω on Σ is a symplectic form on Σ .

If (M, ω) is a symplectic manifold then it is orientable with the standard volume form

$$\Omega_\omega = \frac{(-1)^{\frac{n(n-1)}{2}}}{n!} \omega^n,$$

If $f : M \rightarrow M$ is a symplectic map, then f is volume preserving, $\det_{\Omega_\omega}(f) = 1$ and f is a local diffeomorphism.

In general, if (M, ω) is a $2n$ D compact symplectic manifold then ω^n is a volume element on M , so the *de Rham cohomology class* $[\omega^n] \in H^{2n}(M, \mathbb{R})$ is nonzero. Since $[\omega^n] = [\omega]^n$, $[\omega] \in H^2(M, \mathbb{R})$ and all of its powers through the n th must be nonzero as well. The existence of such an element of

$H^2(M, \mathbb{R})$ is a *necessary condition* for the compact manifold to admit a symplectic structure.

However, if M is a $2n$ D compact manifold without boundary, then there does not exist any exact symplectic structure, $\omega = d\theta$ on M , as its total volume is zero (by Stokes' Theorem),

$$\int_M \Omega_\omega = \frac{(-1)^{\frac{n(n-1)}{2}}}{n!} \int_M \omega^n = \frac{(-1)^{\frac{n(n-1)}{2}}}{n!} \int_M d(\theta \wedge \omega^{n-1}) = 0.$$

For example, spheres S^{2n} do not admit a symplectic structure for $n \geq 2$, since the second de Rham group vanishes, i.e., $H^2(S^{2n}, \mathbb{R}) = 0$. This argument applies to any compact manifold without boundary and having $H^2(M, \mathbb{R}) = 0$.

In mechanics, the phase-space is the cotangent bundle T^*M of a configuration space M . There is a natural symplectic structure on T^*M that is usually defined as follows. Let M be a smooth n D manifold and pick local coordinates $\{dq^1, \dots, dq^n\}$. Then $\{dq^1, \dots, dq^n\}$ defines a basis of the tangent space T_q^*M , and by writing $\theta \in T_q^*M$ as $\theta = p_i dq^i$ we get local coordinates $\{q^1, \dots, q^n, p_1, \dots, p_n\}$ on T^*M . Define the canonical symplectic form ω on T^*M by

$$\omega = dp_i \wedge dq^i.$$

This 2-form ω is obviously independent of the choice of coordinates $\{q^1, \dots, q^n\}$ and independent of the base point $\{q^1, \dots, q^n, p_1, \dots, p_n\} \in T_q^*M$; therefore, it is locally constant, and so $d\omega = 0$.

The canonical 1-form θ on T^*M is the unique 1-form with the property that, for any 1-form β which is a section of T^*M we have $\beta^*\theta = \theta$.

Let $f : M \rightarrow M$ be a diffeomorphism. Then T^*f preserves the canonical 1-form θ on T^*M , i.e., $(T^*f)^*\theta = \theta$. Thus T^*f is symplectic diffeomorphism.

If (M, ω) is a $2n$ D symplectic manifold then about each point $x \in M$ there are local coordinates $\{q^1, \dots, q^n, p_1, \dots, p_n\}$ such that $\omega = dp_i \wedge dq^i$. These coordinates are called canonical or symplectic. By the Darboux Theorem, ω is constant in this local chart, i.e., $d\omega = 0$.

3.12.3 Application: *Autonomous Hamiltonian Mechanics*

3.12.3.1 *Basics of Hamiltonian Mechanics*

In this section we present classical Hamiltonian dynamics. Let (M, ω) be a *symplectic manifold* and $H \in C^\infty(M, \mathbb{R})$ a smooth real valued function on M . The vector-field X_H determined by the condition

$$i_{X_H}\omega + dH = 0,$$

is called *Hamiltonian vector-field* with *Hamiltonian energy function* H . A triple (M, ω, H) is called a *Hamiltonian mechanical system* [Marsden and Ratiu (1999); Puta (1993)].

Nondegeneracy of ω guarantees that X_H exists, but only in the nD case.

Let $\{q^1, \dots, q^n, p_1, \dots, p_n\}$ be *canonical coordinates* on M , i.e., $\omega = dp_i \wedge dq^i$. Then in these coordinates we have

$$X_H = \left(\frac{\partial H}{\partial p_i} \frac{\partial}{\partial q^i} - \frac{\partial H}{\partial q^i} \frac{\partial}{\partial p_i} \right).$$

As a consequence, $((q^i(t)), (p_i(t)))$ is an integral curve of X_H (for $i = 1, \dots, n$) iff *Hamiltonian equations* hold,

$$\dot{q}^i = \partial_{p_i} H, \quad \dot{p}_i = -\partial_{q^i} H. \quad (3.164)$$

Let (M, ω, H) be a Hamiltonian mechanical system and let $\gamma(t)$ be an integral curve of X_H . Then $H(\gamma(t))$ is constant in t . Moreover, if ϕ_t is the flow of X_H , then $H \circ \phi_t = H$ for each t .

Let (M, ω, H) be a Hamiltonian mechanical system and ϕ_t be the flow of X_H . Then, by the Liouville Theorem, for each t , $\phi_t^*\omega = \omega$, ($\frac{d}{dt}\phi_t^*\omega = 0$, so $\phi_t^*\omega$ is constant in t), that is, ϕ_t is symplectic, and it preserves the volume Ω_ω .

A convenient criterion for symplectomorphisms is that they preserve the form of Hamiltonian equations. More precisely, let (M, ω) be a symplectic manifold and $f : M \rightarrow M$ a diffeomorphism. Then f is symplectic iff for all $H \in C^\infty(M, \mathbb{R})$ we have $f^*(X_H) = X_{H \circ f}$.

A vector-field $X \in \mathcal{X}(M)$ on a symplectic manifold (M, ω) is called locally Hamiltonian iff $\mathcal{L}_X\omega = 0$, where L denotes the *Lie derivative*. From the equality $L_{[X, Y]}\omega = \mathcal{L}_X L_Y\omega - L_Y \mathcal{L}_X\omega$, it follows that the locally Hamiltonian vector-fields on M form a Lie subalgebra of $\mathcal{X}(M)$.

Let (M, ω) be a symplectic manifold and $f, g \in C^\infty(M, \mathbb{R})$. The *Poisson*

bracket of f and g is the function

$$\{f, g\}_\omega = -\omega(X_f, X_g) = -\mathcal{L}_{X_f}g = \mathcal{L}_{X_g}f.$$

Also, for $f_0 \in C^\infty(M, \mathbb{R})$, the map $g \mapsto \{f_0, g\}_\omega$ is a derivation. The connection between the Lie bracket and the Poisson bracket is

$$[X_f, X_g] = -X_{\{f, g\}_\omega} \quad \Longleftrightarrow \quad d\omega = 0.$$

The real vector space $C^\infty(M, \mathbb{R})$ together with the Poisson bracket on it forms an infinite-dimensional Lie algebra called the *algebra of classical observables*.

In canonical coordinates $\{q^1, \dots, q^n, p_1, \dots, p_n\}$ on (M, ω) the Poisson bracket of two functions $f, g \in C^\infty(M, \mathbb{R})$ is given by

$$\{f, g\}_\omega = \frac{\partial f}{\partial q^i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q^i}.$$

From this definition follows:

$$\{q^i, q^j\}_\omega = 0, \quad \{p_i, p_j\}_\omega = 0, \quad \{q^i, p_j\}_\omega = \delta_j^i.$$

Let (M, ω) be a symplectic manifold and $f : M \rightarrow M$ a diffeomorphism. Then f is symplectic iff it preserves the Poisson bracket.

Let (M, ω, H) be a Hamiltonian mechanical system and ϕ_t the flow of X_H . Then for each function $f \in C^\infty(M, \mathbb{R})$ we have the equations of motion in the Poisson bracket notation:

$$\frac{d}{dt}(f \circ \phi_t) = \{f \circ \phi_t, H\}_\omega = \{f, H\}_\omega \circ \phi_t.$$

Also, f is called a *constant of motion*, or a *first integral*, if it satisfies the following condition

$$\{f, H\}_\omega = 0.$$

If f and g are constants of motion then their Poisson bracket is also a constant of motion.

A Hamiltonian mechanical system (M, ω, H) is said to be integrable if there exists $n = \frac{1}{2} \dim(M)$ linearly-independent functions $K_1 = H, K_2, \dots, K_n$ such that for each $i, j = 1, 2, \dots, n$:

$$\{K_i, H\}_\omega = 0, \quad \{K_i, K_j\}_\omega = 0.$$

Real 1–DOF Hamiltonian Dynamics

A vector-field $X(t)$ on the *momentum phase-space manifold* M can be given by a system of *canonical equations of motion*

$$\dot{q} = f(q, p, t, \mu), \quad \dot{p} = g(q, p, t, \mu), \quad (3.165)$$

where t is time, μ is a parameter, $q \in S^1$, $p \in \mathbb{R} \times S^1$ are *coordinates* and *momenta*, respectively, while f and g are smooth functions on the phase-space $\mathbb{R} \times S^1$.

If time t does not explicitly appear in the functions f and g , the vector-field X is called *autonomous*. In this case equation (3.165) simplifies as

$$\dot{q} = f(q, p, \mu), \quad \dot{p} = g(q, p, \mu). \quad (3.166)$$

By a *solution curve* of the vector-field X we mean a map $x = (q, p)$, from some interval $I \subset \mathbb{R}$ into the phase-space manifold M , such that $t \mapsto x(t)$. The map $x(t) = (q(t), p(t))$ geometrically represents a curve in M , and equations (3.165) or (3.166) give the tangent vector at each point of the curve.

To specify an *initial condition* on the vector-field X , by

$$x(t, t_0, x_0) = (q(t, t_0, q_0), p(t, t_0, p_0)),$$

geometrically means to distinguish a solution curve by a particular point $x(t_0) = x_0$ in the phase-space manifold M . Similarly, it may be useful to explicitly display the parametric dependence of solution curves, as $x(t, t_0, x_0, \mu) = (q(t, t_0, q_0, \mu_q), p(t, t_0, p_0, \mu_p))$, where μ_q, μ_p denote q -dependent and p -dependent parameters, respectively.

The solution curve $x(t, t_0, x_0)$ of the vector-field X , may be also referred as the *phase trajectory* through the point x_0 at $t = t_0$. Its graph over t is referred to as an *integral curve*; more precisely, *graph*

$$x(t, t_0, x_0) \equiv \{(x, t) \in M \times \mathbb{R} : x = x(t, t_0, x_0), t \in I \subset \mathbb{R}\}.$$

Let $x_0 = (q_0, p_0)$ be a point on M . By the *orbit through* x_0 , denoted $O(x_0)$, we mean the set of points in M that lie on a trajectory passing through x_0 ; more precisely, for $x_0 \in U$, U open in M , the orbit through x_0 is given by $O(x_0) = \{x \in \mathbb{R} \times S^1 : x = x(t, t_0, x_0), t \in I \subset \mathbb{R}\}$.

Consider a general autonomous vector-field X on the phase-space manifold M , given by equation $\dot{x} = f(x)$, $x = (q, p) \in M$. An *equilibrium solution*, *singularity*, or *fixed point* of X is a point $\bar{x} \in M$ such that $f(\bar{x}) = 0$, i.e., a solution which does not change in time.

Any solution $\bar{x}(t)$ of an autonomous vector-field X on M is *stable* if solutions starting 'close' to $\bar{x}(t)$ at a given time remain close to $\bar{x}(t)$ for all later times. It is *asymptotically stable* if nearby solutions actually converge to $\bar{x}(t)$ as $t \rightarrow \infty$. In order to determine the stability of $\bar{x}(t)$ we must understand the nature of solutions near $\bar{x}(t)$, which is done by *linearization* of the vector-field X . The solution of the linearized vector-field Y is asymptotically stable if all eigenvalues have negative real parts. In that case the fixed point $x = \bar{x}$ of associated nonlinear vector-field X is also asymptotically stable. A fixed point \bar{x} is called *hyperbolic point* if none of the eigenvalues of Y have zero real part; in that case the orbit structure near \bar{x} is essentially the same for X and Y .

In the case of autonomous vector-fields on M we have also an important property of *Hamiltonian flow*. If $x(t) = (q(t), p(t))$ is a solution of $\dot{x} = f(x)$, $x \in M$, then so is $x(t + \tau)$ for any $\tau \in \mathbb{R}$. Also, for any $x_0 \in M$ there exists only one solution of an autonomous vector-field passing through this point. The autonomous vector-field

$$\dot{x} = f(x)$$

has the following properties (compare with the section (6.289) above):

- (1) $x(t, x_0)$ is C^∞ ;
- (2) $x(0, x_0) = x_0$; and
- (3) $x(t + s, x_0) = x(t, x(s, x_0))$.

These properties show that the solutions of an autonomous vector-field form a *one-parameter group of diffeomorphisms* of the phase-space manifold M . This is referred to as a *phase-flow* and denoted by $\phi_t(x)$ or $\phi(t, x)$.

Consider a flow $\phi(t, x)$ generated by vector-field $\dot{x} = f(x)$. A point $x_0 = (q_0, p_0)$ on M is called an ω -*limit point* of $x, = (q, p) \in M$, denoted $\omega(x)$, if there exists a sequence $\{t_i\}$, $t_i \mapsto \infty$, such that $\phi(t_i, x) \mapsto x_0$. Similarly, α -*limit points* are defined by taking a sequence $\{t_i\}$, $t_i \mapsto -\infty$. The set of all ω -limit points of a flow is called the ω -*limit set*. The α -limit set is similarly defined.

A point $x_0 = (q_0, p_0)$ on M is called *nonwandering* if for any open neighborhood $U \subset M$ of x_0 , there exists some $t \neq 0$ such that $\phi(t, U) \cap U \neq \emptyset$. The set of all nonwandering points of a flow is called the *nonwandering set* of that particular map or flow.

A closed invariant subset $A \subset M$ is called an *attracting set* if there is some open neighborhood $U \subset M$ of A such that $\phi(t, x) \in U$ and $\phi(t, x) \mapsto A$ for any $x \in U$ and $t \geq 0$. The *domain* or *basin of attraction* of A is given

by $\cup_{t \leq 0} \phi(t, U)$. In practice, a way of locating attracting sets is to first find a *trapping region*, i.e., a closed, connected subset $V \subset M$ such that for any $t \geq 0$ $\phi(t, V) \subset V$. Then $\cap_{t \geq 0} \phi(t, V) = A$ is an *attracting set*.

As a first example of one-DOF dynamical systems, let us consider a vector-field $x = (q, p) \in \mathbb{R} \times \mathbb{R}$ of a simple harmonic oscillator, given by equations

$$\dot{q} = p, \quad \dot{p} = -q. \quad (3.167)$$

Here, the *solution* passing through the point $(q, p) = (1, 0)$ at $t = 0$ is given by $(q(t), p(t)) = (\cos t, -\sin t)$; the *integral curve* passing through $(q, p) = (1, 0)$ at $t = 0$ is given by $\{(q, p, t) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R} : (q(t), p(t)) = (\cos t, -\sin t)\}$, for all $t \in \mathbb{R}$; the *orbit* passing through $(q, p) = (1, 0)$ is given by the circle $q^2 + p^2 = 1$.

A one-DOF dynamical system is called *Hamiltonian system* if there exists a *first integral* or a function of the dependent variables (q, p) whose level curves give the orbits of the vector-field $X = X_H$, i.e., a total-energy *Hamiltonian function* $H = H(q, p) : U \rightarrow \mathbb{R}$, (U open set on the phase-space manifold M), such that the vector-field X_H is given by *Hamiltonian canonical equations* (3.164). In (3.164), the first, \dot{q} -equation, is called the *velocity equation* and serves as a definition of the *momentum*, while the second, \dot{p} -equation is called the *force equation*, and represents the *Newtonian second law of motion*.

The simple harmonic oscillator (3.167) is a Hamiltonian system with a Hamiltonian function $H = \frac{p^2}{2} + \frac{q^2}{2}$. It has a *fixed point* – *center* (having purely imaginary eigenvalues) at $(q, p) = (0, 0)$ and is surrounded by a one-parameter family of periodic orbits given by the Hamiltonian H .

A nice example of one-DOF dynamical system with a Hamiltonian structure is a *damped Duffing oscillator* (see, e.g., [Wiggins (1990)]). This is a *plane* Hamiltonian vector-field $x = (q, p) \in \mathbb{R}^2$, given by Hamiltonian equations

$$\dot{q} = p \equiv f(q, p), \quad \dot{p} = q - q^3 - \delta p \equiv g(q, p, \delta), \quad \delta \geq 0. \quad (3.168)$$

For the special parameter value $\delta = 0$, we have an *undamped* Duffing oscillator with a *first integral* represented by Hamiltonian function $H = \frac{p^2}{2} - \frac{q^2}{2} + \frac{q^4}{4}$, where $\frac{p^2}{2}$ corresponds to the *kinetic energy* (with a mass scaled to unity), and $-\frac{q^2}{2} + \frac{q^4}{4} \equiv V(x)$ corresponds to the *potential energy* of the oscillator.

In general, if the first integral, i.e., a Hamiltonian function H , is defined by $H = \frac{p^2}{2} + V(x)$, then the momentum is given by $p = \pm \sqrt{2(H - V(x))}$.

All one-DOF Hamiltonian systems are *integrable* and all the solutions lie on *level curves* of the Hamiltonian function, which are topologically equivalent with the circle S^1 . This is actually a general characteristic of all n -DOF integrable Hamiltonian systems: their bounded motions lie on nD *invariant tori* $T^n = S^1 \times \cdots \times S^1$, or *homoclinic orbits*. The homoclinic orbit is sometimes called a *separatrix* because it is the boundary between two distinctly different types of motion.

For example, in case of a damped Duffing oscillator (3.168) with $\delta \neq 0$, we have

$$\partial_q f + \partial_p g = -\delta,$$

and according to the *Bendixon criterion* for $\delta > 0$ it has no closed orbits.

The vector-field X given by equations (3.168) has three fixed points given by $(q, p) = (0, 0), (\pm 1, 0)$. The *eigenvalues* $\lambda_{1,2}$ of the associated linearized vector-field are given by $\lambda_{1,2} = -\delta/2 \pm \frac{1}{2}\sqrt{\delta^2 + 4}$, for the fixed point $(0, 0)$, and by $\lambda_{1,2} = -\delta/2 \pm \frac{1}{2}\sqrt{\delta^2 - 8}$, for the fixed point $(\pm 1, 0)$. Hence, for $\delta > 0$, $(0, 0)$ is *unstable* and $(\pm 1, 0)$ are *asymptotically stable*; for $\delta = 0$, $(\pm 1, 0)$ are *stable in the linear approximation* (see, e.g., [Wiggins (1990)]).

Another example of one-DOF Hamiltonian systems is a *simple pendulum* (again, all physical constants are scaled to unity), given by Hamiltonian function $H = \frac{p^2}{2} - \cos q$. This is the first integral of the *cylindrical* Hamiltonian vector-field $(q, p) \in S^1 \times \mathbb{R}$, defined by canonical equations

$$\dot{q} = p, \quad \dot{p} = -\sin q.$$

This vector-field has fixed points at $(0, 0)$, which is a center (i.e., the eigenvalues are purely imaginary), and at $(\pm\pi, 0)$, which are saddles, but since the phase-space manifold is the cylinder, these are really the same point.

The basis of human arm and leg dynamics represents the *coupling* of two uniaxial, $SO(2)$ -joints. The study of two DOF Hamiltonian dynamics we shall start with the most simple case of two linearly coupled linear undamped oscillators with parameters scaled to unity. Under general conditions we can perform a change of variables to canonical coordinates (the ‘normal modes’) (q^i, p_i) , $i = 1, 2$, so that the vector-field X_H is given by

$$\dot{q}^1 = p_1, \quad \dot{q}^2 = p_2, \quad \dot{p}_1 = -\omega_1^2 q^1, \quad \dot{p}_2 = -\omega_2^2 q^2.$$

This system is integrable, since we have two independent functions of

(q^i, p_i) , i.e., Hamiltonians

$$H_1 = \frac{p_1^2}{2} + \frac{\omega_1^2 (q^1)^2}{2}, \quad H_2 = \frac{p_2^2}{2} + \frac{\omega_2^2 (q^2)^2}{2}.$$

The level curves of these functions are compact sets (topological circles); therefore, the orbits in the 4D phase-space \mathbb{R}^4 actually lie on the two-torus T^2 . By making the appropriate change of variables, it can be shown (see, e.g., [Wiggins (1990)]) that the whole dynamics of the two linearly coupled linear undamped oscillators is actually contained in the equations

$$\dot{\theta}_1 = \omega_1, \quad \dot{\theta}_2 = \omega_2, \quad (\theta_1, \theta_2) \in S^1 \times S^1 \equiv T^2. \quad (3.169)$$

The flow on the two-torus T^2 , generated by (3.169), is simple to calculate and is given by

$$\theta_1(t) = \omega_1 t + \theta_{10}, \quad \theta_2(t) = \omega_2 t + \theta_{20}, \quad (\text{mod } 2\pi),$$

and θ_1 and θ_2 are called the longitude and latitude. However, orbits under this flow will depend on how ω_1 and ω_2 are related. If ω_1 and ω_2 are *commensurate* (i.e., the equation $m\omega_1 + n\omega_2 = 0$, $(n, m) \in \mathbb{Z}$ has solutions), then every phase curve of (3.169) is closed. However, if ω_1 and ω_2 are *incommensurate* i.e., upper equation has no solutions), then every phase curve of (3.169) is everywhere dense on T^2 .

Somewhat deeper understanding of Hamiltonian dynamics is related to the method of *action-angle variables*. The easiest way to introduce this idea is to consider again a simple harmonic oscillator (3.167). If we transform equations (3.167) into polar coordinates using $q = r \sin \theta$, $p = r \cos \theta$, then the equations of the vector-field become $\dot{r} = 0$, $\dot{\theta} = 1$, having the obvious solution $r = \text{const}$, $\theta = t + \theta_0$. For this example polar coordinates work nicely because the system (3.167) is linear and, therefore, all of the periodic orbits have the same period.

For the general, nonlinear one-DOF Hamiltonian system (3.164) we will seek a coordinate transformation that has the same effect. Namely, we will seek a coordinate transformation $(q, p) \mapsto (\theta(q, p), I(q, p))$ with inverse transformation $(\theta, I) \mapsto (q(I, \theta), p(I, \theta))$ such that the vector-field (3.164) in the action-angle (θ, I) coordinates satisfies the following conditions: (i) $\dot{I} = 0$; (ii) θ changes linearly in time on the closed orbits with $\dot{\theta} = \Omega(I)$. We might even think of I and θ heuristically as 'nonlinear polar coordinates'. In such a coordinate system Hamiltonian function takes the form $H = H(I)$, and also, $\Omega(I) = \partial_I H$, i.e., specifying I specifies a periodic orbit.

The action variable $I(q, p)$ geometrically represents an area enclosed by any closed curve, which is constant in time. It is defined as an integral $I = \frac{1}{2\pi} \int_H p dq$, where H denotes the periodic orbit defined by $H(q, p) = H = \text{const}$. If the period of each periodic orbit defined by $H(q, p) = H = \text{const}$ is denoted by $T(H)$, the angle variable $\theta(q, p)$ is defined by

$$\theta(q, p) = \frac{2\pi}{T(H)} t(q, p),$$

where $t = t(q, p)$ represents the time taken for the solution starting from (q_0, p_0) to reach (q, p) .

For the system with Hamiltonian $H = \frac{p^2}{2} + V(x)$ and momentum $p = \pm\sqrt{2}\sqrt{H - V(x)}$ the action is given by $I = \frac{\sqrt{2}}{\pi} \int_{q_{\min}}^{q_{\max}} \sqrt{H - V(q)} dq$, and the angle is given by $\theta(q, p) = \frac{2\pi}{T(H)} \int_{q_{\min}}^{q_{\max}} \frac{dq}{\sqrt{2}\sqrt{H - V(q)}}$.

Closely related to the action-angle variables is the *perturbation theory* (see [Nayfeh (1973)]). To explain the main idea of this theory, let us consider an ϵ -perturbed vector-field periodic in t which can be in component form given as (with $(q, p) \in \mathbb{R}^2$)

$$\dot{q} = f_1(q, p) + \epsilon g_1(q, p, t, \epsilon), \quad \dot{p} = f_2(q, p) + \epsilon g_2(q, p, t, \epsilon). \quad (3.170)$$

Setting $\epsilon = 0$ we get the *unperturbed* Hamiltonian system with a smooth scalar-valued function $H(q, p)$ for which holds

$$f_1(q, p) = \frac{\partial H(q, p)}{\partial p}, \quad f_2(q, p) = -\frac{\partial H(q, p)}{\partial q},$$

so, the perturbed system (3.170) gets the symmetric canonical form

$$\dot{q} = \frac{\partial H(q, p)}{\partial p} + \epsilon g_1(q, p, t, \epsilon), \quad \dot{p} = -\frac{\partial H(q, p)}{\partial q} + \epsilon g_2(q, p, t, \epsilon).$$

The perturbation (g_1, g_2) need not be Hamiltonian, although in the case where perturbation is Hamiltonian versus the case where it is not, the dynamics are very different.

Now, if we transform the coordinates of the perturbed vector-field using the action-angle transformation for the unperturbed Hamiltonian vector-field, we get

$$\dot{I} = \epsilon \left(\frac{\partial I}{\partial q} g_1 + \frac{\partial I}{\partial p} g_2 \right) \equiv \epsilon F(I, \theta, t, \epsilon), \quad (3.171)$$

$$\dot{\theta} = \Omega(I) + \epsilon \left(\frac{\partial \theta}{\partial q} g_1 + \frac{\partial \theta}{\partial p} g_2 \right) \equiv \Omega(I) + \epsilon G(I, \theta, t, \epsilon), \quad \text{where}$$

$$\begin{aligned}
F(I, \theta, t, \epsilon) &= \frac{\partial I}{\partial q}(q(I, \theta), p(I, \theta)) g_1((q(I, \theta), p(I, \theta), t, \epsilon) \\
&\quad + \frac{\partial I}{\partial p}(q(I, \theta), p(I, \theta)) g_2((q(I, \theta), p(I, \theta), t, \epsilon), \quad \text{and} \\
G(I, \theta, t, \epsilon) &= \frac{\partial \theta}{\partial q}(q(I, \theta), p(I, \theta)) g_1((q(I, \theta), p(I, \theta), t, \epsilon) \\
&\quad + \frac{\partial \theta}{\partial p}(q(I, \theta), p(I, \theta)) g_2((q(I, \theta), p(I, \theta), t, \epsilon).
\end{aligned}$$

Here, F and G are 2π periodic in θ and $T = 2\pi/\omega$ periodic in t .

Finally, we shall explain in brief the most important idea in the dynamical systems theory, the idea of *Poincaré maps*. The idea of reducing the study of continuous time systems (flows) to the study of an *associated discrete time system (map)* is due to Poincaré who first utilized it in the end of the last Century in his studies of the three body problem in celestial mechanics. Nowadays virtually any discrete time system that is associated with an ordinary differential equation is referred to as a Poincaré map [Wiggins (1990)]. This technique offers several advantages in the study of dynamical systems, including dimensional reduction, global dynamics and conceptual clarity. However, construction of a Poincaré map requires some knowledge of the phase-space of a dynamical system. One of the techniques which can be used for construction of Poincaré maps is the perturbation method.

To construct the Poincaré map for the system (3.171), we have to rewrite it as an autonomous system

$$\dot{I} = \epsilon F(I, \theta, \phi, \epsilon), \quad \dot{\theta} = \Omega(I) + \epsilon G(I, \theta, \phi, \epsilon), \quad \dot{\phi} = \omega, \quad (3.172)$$

(where $(I, \theta, \phi) \in \mathbb{R}^+ \times S^1 \times S^1$. We construct a global *cross-section* Σ to this vector-field defined as $\Sigma^{\phi_0} = \{(I, \theta, \phi) | \phi = \phi_0\}$. If we denote the (I, θ) components of solutions of (3.172) by $(I_\epsilon(t), \theta_\epsilon(t))$ and the (I, θ) components of solutions of (3.172) for $\epsilon = 0$ by $(I_0, \Omega(I_0)t + \theta_0)$, then the perturbed Poincaré map is given by

$$P_\epsilon : \Sigma^{\phi_0} \rightarrow \Sigma^{\phi_0}, \quad (I_\epsilon(0), \theta_\epsilon(0)) \mapsto (I_\epsilon(T), \theta_\epsilon(T)),$$

and the m th *iterate* of the Poincaré map is given by

$$P_\epsilon^m : \Sigma^{\phi_0} \rightarrow \Sigma^{\phi_0}, \quad (I_\epsilon(0), \theta_\epsilon(0)) \mapsto (I_\epsilon(mT), \theta_\epsilon(mT)).$$

Now we can approximate the solutions to the perturbed problem as

linear, constant-coefficient approximation

$$I_\epsilon(t) = I_0 + \epsilon I_1(t) + O(\epsilon^2), \quad \theta_\epsilon(t) = \theta_0 + \Omega(I_0)t + \epsilon \theta_1(t) + O(\epsilon^2),$$

where we have chosen $I_\epsilon(0) = I_0$, $\theta_\epsilon(0) = \theta_0$.

As a last example of one-DOF Hamiltonian dynamics we shall analyze a *damped, forced Duffing oscillator*, given by canonical equations [Wiggins (1990)]

$$\dot{q} = p, \quad \dot{p} = q - q^3 - \delta p + \gamma \cos \omega t, \quad \delta, \gamma, \omega \geq 0, \quad (q, p) \in \mathbb{R}^2. \quad (3.173)$$

where δ, γ , and ω are real parameters physically meaning *dissipation*, *amplitude of forcing* and *frequency*, respectively.

The *perturbed* system (3.173) is given by

$$\dot{q} = p, \quad \dot{p} = q - q^3 + \epsilon(\gamma \cos \omega t - \delta p), \quad (3.174)$$

where ϵ -perturbation is assumed small. Then the *unperturbed* system reads

$$\dot{q} = p, \quad \dot{p} = q - q^3.$$

It is conservative with Hamiltonian function

$$H(q, p) = \frac{p^2}{2} - \frac{q^2}{2} + \frac{q^4}{4}. \quad (3.175)$$

In the unperturbed phase-space all orbits are given by the level sets of the Hamiltonian (3.175). There are three equilibrium points at the following coordinates: $(q, p) = (\pm 1, 0)$ – centers, and $(q, p) = (0, 0)$ – saddle. The saddle point is connected to itself by two homoclinic orbits given by

$$q_+^0(t) = (\sqrt{2}(\cosh t)^{-1}, -\sqrt{2}(\cosh t)^{-1} \tanh t), \quad q_-^0(t) = -q_+^0(t).$$

There are two families of *periodic orbits* $q_\pm^k(t)$, where k represents the *elliptic modulus* related to the Hamiltonian by $H(q_\pm^k(t)) \equiv H(k) = \frac{k^2-1}{(2-k^2)^2}$, inside the corresponding homoclinic orbits $q_\pm^0(t)$, with the period $T(k) = 2K(k)\sqrt{2-k^2}$ ($K(k)$ is the complete elliptic integral of the first kind).

Also, there exists a family of periodic orbits outside the homoclinic orbits with the period $T(k) = 4K(k)\sqrt{k^2-1}$.

The perturbed system (3.174) can be rewritten as a third-order autonomous system

$$\dot{q} = p, \quad \dot{p} = q - q^3 + \epsilon(\gamma \cos \phi - \delta p), \quad \dot{\phi} = \omega,$$

where $(q, p, \phi) \in \mathbb{R}^2 \times S^1$, S^1 is the circle of length $2\pi/\omega$ and $\phi(t) = \omega t + \phi_0$. We form the global cross-section to the flow

$$\Sigma^{\phi_0} = \{(q, p, \phi) | \phi = \phi_0 \in [0, 2\pi/\omega]\}$$

and the associated Poincaré map is given by

$$P : \Sigma^{\phi_0} \rightarrow \Sigma^{\phi_0}, \quad (q(0), p(0)) \mapsto (q(2\pi/\omega), p(2\pi/\omega)).$$

A detailed analysis of the perturbed Poincaré map for the damped, forced Duffing oscillator is related to the *Melnikov function* (see [Wiggins (1990)]).

Complex 1-DOF Hamiltonian Dynamics

Recall that setting $z = q + ip$, $z \in \mathbb{C}$, $i = \sqrt{-1}$, Hamiltonian equations $\dot{q} = \partial H / \partial p$, $\dot{p} = -\partial H / \partial q$ may be written in *complex notation* as [Abraham and Marsden (1978); Marsden and Ratiu (1999); Wiggins (1990)]

$$\dot{z} = -2i \frac{\partial H}{\partial \bar{z}}. \quad (3.176)$$

Let U be an open set in the *complex phase-space manifold* $M_{\mathbb{C}}$ (i.e., manifold M modelled on \mathbb{C}). A C^0 function $\gamma : [a, b] \rightarrow U \subset M_{\mathbb{C}}$, $t \mapsto \gamma(t)$ represents a *solution curve* $\gamma(t) = q(t) + ip(t)$ of a complex Hamiltonian system (3.176). For example, the curve $\gamma(\theta) = \cos \theta + i \sin \theta$, $0 \leq \theta \leq 2\pi$ is the unit circle. $\gamma(t)$ is a *parameterized curve*. We call $\gamma(a)$ the *beginning point*, and $\gamma(b)$ the *end point* of the curve. By a *point on the curve* we mean a point w such that $w = \gamma(t)$ for some $t \in [a, b]$.

The derivative $\dot{\gamma}(t)$ is defined in the usual way, namely

$$\dot{\gamma}(t) = \dot{q}(t) + i\dot{p}(t),$$

so that the usual rules for the derivative of a sum, product, quotient, and chain rule are valid. The *speed* is defined as usual to be $|\dot{\gamma}(t)|$. Also, if $f : U \rightarrow M_{\mathbb{C}}$ represents a *holomorphic*, or *analytic* function, then the composite $f \circ \gamma$ is differentiable (as a function of the real variable t) and $(f \circ \gamma)'(t) = f'(\gamma(t)) \dot{\gamma}(t)$.

Recall that a *path* represents a sequence of C^1 -curves,

$$\gamma = \{\gamma_1, \gamma_2, \dots, \gamma_n\},$$

such that the end point of γ_j , ($j = 1, \dots, n$) is equal to the beginning point of γ_{j+1} . If γ_j is defined on the interval $[a_j, b_j]$, this means that

$$\gamma_j(b_j) = \gamma_{j+1}(a_{j+1}).$$

We call $\gamma_1(a_1)$ the *beginning point* of γ_j , and $\gamma_n(b_n)$ the *end point* of γ_j . The path is said to *lie in an open set* $U \subset M_C$ if each curve γ_j lies in U , i.e., for each t , the point $\gamma_j(t)$ lies in U .

An open set U is *connected* if given two points α and β in U , there exists a path $\gamma = \gamma_1, \gamma_2, \dots, \gamma_n$ in U such that α is the beginning point of γ_1 and β is the end point of γ_n ; in other words, if there is a path γ in U which joins α to β . If U is a connected open set and f a holomorphic function on U such that $f' = 0$, then f is a constant. If g is a function on U such that $f' = g$, then f is called a *primitive* of g on U . Primitives can be either find out by integration or written down directly.

Let f be a C^0 -function on an open set U , and suppose that γ is a curve in U , meaning that all values $\gamma(t)$ lie in U for $a \leq t \leq b$. The *integral of f along γ* is defined as

$$\int_{\gamma} f = \int_{\gamma} f(z) = \int_a^b f(\gamma(t)) \dot{\gamma}(t) dt.$$

For example, let $f(z) = 1/z$, and $\gamma(\theta) = e^{i\theta}$. Then $\dot{\gamma}(\theta) = ie^{i\theta}$. We want to find the value of the integral of f over the circle, $\int_{\gamma} dz/z$, so $0 \leq \theta \leq 2\pi$. By definition, this integral is equal to $\int_0^{2\pi} ie^{i\theta}/e^{i\theta} d\theta = i \int_0^{2\pi} d\theta = 2\pi i$.

The *length* $L(\gamma)$ is defined to be the integral of the speed, $L(\gamma) = \int_a^b |\dot{\gamma}(t)| dt$.

If $\gamma = \gamma_1, \gamma_2, \dots, \gamma_n$ is a path, then the integral of a C^0 -function f on an open set U is defined as $\int_{\gamma} f = \sum_{i=1}^n \int_{\gamma_i} f$, i.e., the sum of the integrals of f over each curve γ_i ($i = 1, \dots, n$) of the path γ . The *length of a path* is defined as $L(\gamma) = \sum_{i=1}^n L(\gamma_i)$.

Let f be continuous on an open set $U \subset M_C$, and suppose that f has a primitive g , that is, g is holomorphic and $g' = f$. Let α, β be two points in U , and let γ be a path in U joining α to β . Then $\int_{\gamma} f = g(\beta) - g(\alpha)$; this integral is independent of the path and depends only on the beginning and end point of the path.

A *closed path* is a path whose beginning point is equal to its end point. If f is a C^0 -function on an open set $U \subset M_C$ admitting a holomorphic primitive g , and γ is any closed path in U , then $\int_{\gamma} f = 0$.

Let γ, η be two paths defined over the same interval $[a, b]$ in an open set

$U \subset M_C$. Recall (see Introduction) that γ is *homotopic* to η if there exists a C^0 -function $\psi : [a, b] \times [c, d] \rightarrow U$ defined on a rectangle $[a, b] \times [c, d] \subset U$, such that $\psi(t, c) = \gamma(t)$ and $\psi(t, d) = \eta(t)$ for all $t \in [a, b]$. For each number $s \in [c, d]$ we may view the function $\psi_s(t) = \psi(t, s)$ as a continuous curve defined on $[a, b]$, and we may view the family of continuous curves ψ_s as a *deformation* of the path γ to the path η . It is said that the homotopy ψ *leaves the end points fixed* if we have $\psi(a, s) = \gamma(a)$ and $\psi(b, s) = \gamma(b)$ for all values of $s \in [c, d]$. Similarly, when we speak of a homotopy of closed paths, we assume that each path ψ_s is a closed path.

Let γ, η be paths in an open set $U \subset M_C$ having the same beginning and end points. Assume that they are homotopic in U . Let f be holomorphic on U . Then $\int_\gamma f = \int_\eta f$. The same holds for closed homotopic paths in U . In particular, if γ is homotopic to a point in U , then $\int_\gamma f = 0$. Also, it is said that an open set $U \subset M_C$ is *simply-connected* if it is connected and if every closed path in U is homotopic to a point.

In the previous example we found that

$$\frac{1}{2\pi i} \int_\gamma \frac{1}{z} dz = 1,$$

if γ is a circle around the origin, oriented counterclockwise. Now we define for any closed path γ its *winding number* with respect to a point α to be

$$W(\gamma, \alpha) = \frac{1}{2\pi i} \int_\gamma \frac{1}{z - \alpha} dz,$$

provided the path does not pass through α . If γ is a closed path, then $W(\gamma, \alpha)$ is an integer.

A closed path $\gamma \in U \subset M_C$ is *homologous to 0 in U* if

$$\int_\gamma \frac{1}{z - \alpha} dz = 0,$$

for every point α not in U , or in other words, $W(\gamma, \alpha) = 0$ for every such point.

Similarly, let γ, η be closed paths in an open set $U \subset M_C$. We say that they are *homologous in U*, and write $\gamma \sim \eta$, if $W(\gamma, \alpha) = W(\eta, \alpha)$ for every point α in the complement of U . We say that γ is *homologous to 0 in U*, and write $\gamma \sim 0$, if $W(\gamma, \alpha) = 0$ for every point α in the complement of U .

If γ and η are closed paths in U and are homotopic, then they are homologous. If γ and η are closed paths in U and are close together, then they are homologous.

Let $\gamma_1, \dots, \gamma_n$ be curves in an open set $U \subset M_C$, and let m_1, \dots, m_n be integers. A formal sum $\gamma = m_1\gamma_1 + \dots + m_n\gamma_n = \sum_{i=1}^n m_i\gamma_i$ is called a *chain in U* . The chain is called *closed* if it is a finite sum of closed paths. If γ is the chain as above, then $\int_\gamma f = \sum_i m_i \int_{\gamma_i} f$. If γ and η are closed chains in U , then $W(\gamma + \eta, \alpha) = W(\gamma, \alpha) + W(\eta, \alpha)$. We say that γ and η are *homologous in U* , and write $\gamma \sim \eta$, if $W(\gamma, \alpha) = W(\eta, \alpha)$ for every point α in the complement of U . We say that γ is *homologous to 0 in U* , and write $\gamma \sim 0$, if $W(\gamma, \alpha) = 0$ for every point α in the complement of U .

Recall that the *Cauchy Theorem* states that if γ is a closed chain in an open set $U \subset M_C$, and γ is homologous to 0 in U , then $\int_\gamma f = 0$. If γ and η are closed chains in U , and $\gamma \sim \eta$ in U , then $\int_\gamma f = \int_\eta f$.

It follows from Cauchy's Theorem that if γ and η are homologous, then $\int_\gamma f = \int_\eta f$ for all holomorphic functions f on U [Abraham and Marsden (1978); Wiggins (1990)].

3.12.3.2 Library of Basic Hamiltonian Systems

In this subsection, we present some basic Hamiltonian systems used by human-like biodynamics (for more details, see [Putz (1993)]).

1D Harmonic Oscillator

In this case we have $\{p, q\}$ as canonical coordinates on \mathbb{R}^2

$$\begin{aligned} M &= T^*\mathbb{R} \simeq \mathbb{R}^2, & \omega &= dp \wedge dq, \\ H &= \frac{1}{2}(p^2 + q^2), & X_H &= p \frac{\partial}{\partial q} - q \frac{\partial}{\partial p}, \end{aligned}$$

and Hamiltonian equations read

$$\dot{q} = p, \quad \dot{p} = -q.$$

For each $f, g \in C^\infty(\mathbb{R}^2, \mathbb{R})$ the Poisson bracket is given by

$$\{f, g\}_\omega = \frac{\partial f}{\partial q} \frac{\partial g}{\partial p} - \frac{\partial f}{\partial p} \frac{\partial g}{\partial q}.$$

Complex Plane

Let $T^*\mathbb{R} \simeq \mathbb{R}^2$ have the canonical symplectic structure $\omega = dp \wedge dq$.

Writing $z = q + ip$, we have

$$\omega = \frac{1}{2i} dz \wedge d\bar{z}, \quad X_H = i \left(\frac{\partial H}{\partial z} \frac{\partial}{\partial z} - \frac{\partial H}{\partial \bar{z}} \frac{\partial}{\partial \bar{z}} \right),$$

$$\{f, g\}_\omega = \frac{i}{2} \left(\frac{\partial f}{\partial z} \frac{\partial g}{\partial \bar{z}} - \frac{\partial f}{\partial \bar{z}} \frac{\partial g}{\partial z} \right),$$

so, the Hamiltonian equations, $\dot{q} = \partial_p H$, $\dot{p} = -\partial_q H$, become

$$\dot{z} = -2i \frac{\partial H}{\partial \bar{z}}.$$

2D Harmonic Oscillator

In this case we have $\{q^1, y, p_1, p_2\}$ as canonical coordinates on \mathbb{R}^4

$$M = T^*\mathbb{R}^2 \simeq \mathbb{R}^4, \quad \omega = dp_1 \wedge dq^1 + dp_2 \wedge dq^2,$$

$$H = \frac{1}{2} [p_1^2 + p_2^2 + (q^1)^2 + (y)^2].$$

The functions $f = p_i p_j + q^i q^j$ and $g = p_i q^j + p_j q^i$, (for $i, j = 1, 2$), are constants of motion.

n D Harmonic Oscillator

In this case we have ($i = 1, \dots, n$)

$$M = T^*\mathbb{R}^n \simeq \mathbb{R}^{2n}, \quad \omega = dp_i \wedge dq^i,$$

$$H = \frac{1}{2} \sum_{i=1}^n [p_i^2 + (q^i)^2].$$

The system is integrable in an open set of $T^*\mathbb{R}^n$ with:

$$K_1 = H, \quad K_2 = p_2^2 + (y)^2, \quad \dots, \quad K_n = p_n^2 + (q^n)^2.$$

Toda Molecule

Consider three mass-points on the line with coordinates q^i , ($i = 1, 2, 3$), and satisfying the ODEs:

$$\ddot{q}^i = -\partial_{q^i} U, \quad \text{where} \quad U = e^{q^1 - q^2} + e^{q^2 - q^3} - e^{q^3 - q^1}.$$

This is a Hamiltonian system with $\{q^i, p_i\}$ as canonical coordinates on \mathbb{R}^6 ,

$$M = T^*\mathbb{R}^3 \simeq \mathbb{R}^6, \quad \omega = dp_i \wedge dq^i,$$

$$H = \frac{1}{2} (p_1^2 + p_2^2 + p_3^2) + U.$$

The *Toda molecule* (3.12.3.2) is an integrable Hamiltonian system in an open set of $T^*\mathbb{R}^3$ with:

$$K_1 = H, \quad K_2 = p_1 + p_2 + p_3,$$

$$K_3 = \frac{1}{9} (p_1 + p_2 + p_3) (p_2 + p_3 - 2p_1) (p_3 + p_1 - 2p_2) - (p_1 + p_2 - 2p_3) e^{q^1 - q^2}$$

$$- (p_2 + p_3 - 2p_1) e^{q^2 - q^3} - (p_3 + p_1 - 2p_2) e^{q^3 - q^1}.$$

3-Point Vortex Problem

The motion of three-point vortices for an ideal incompressible fluid in the plane is given by the equations:

$$\dot{q}^j = -\frac{1}{2\pi} \sum_{i \neq j} \Gamma_i (p_j - p_i) / r_{ij}^2,$$

$$\dot{p}_j = \frac{1}{2\pi} \sum_{i \neq j} \Gamma_i (q^i - q^j) / r_{ij}^2,$$

$$r_{ij}^2 = (q^i - q^j)^2 + (p_j - p_i)^2,$$

where $i, j = 1, 2, 3$, and Γ_i are three nonzero constants. This mechanical system is Hamiltonian if we take:

$$M = T^*\mathbb{R}^3 \simeq \mathbb{R}^6, \quad \omega = dp_i \wedge dq^i, \quad (i = 1, \dots, 3),$$

$$H = -\frac{1}{4\pi} \sum_{i,j=1}^3 \Gamma_i \Gamma_j \ln(r_{ij}).$$

Moreover, it is integrable in an open set of $T^*\mathbb{R}^3$ with:

$$K_1 = H, \quad K_2 = \sum_{i=1}^3 \Gamma_i \left[(q^i)^2 + p_i^2 \right],$$

$$K_3 = \left(\sum_{i=1}^3 \Gamma_i q^i \right)^2 + K_2^2.$$

The Newton's Second Law as a Hamiltonian System

In the case of conservative forces, Newton's law of motion can be written on \mathbb{R}^{3n} as

$$m_i \ddot{q}^i = -\partial_{q^i} U, \quad (i = 1, 2, \dots, 3n).$$

Its symplectic formulation reads:

$$M = T^*\mathbb{R}^3 \simeq \mathbb{R}^6, \quad \omega = dp_i \wedge dq^i,$$

$$H = \sum_{i=1}^{3n} \frac{p_i^2}{2m_i} + U.$$

The Hamiltonian vector-field X_H is

$$X_H = \left(\frac{p_i}{m_i} \partial_{q^i} - \partial_{q^i} U \partial_{p_i} \right),$$

giving the Hamiltonian equations

$$\dot{q}^i = \frac{p_i}{m_i}, \quad \dot{p}_i = -\partial_{q^i} U.$$

Rigid Body Fixed in a Point

The configuration space of a *rigid body fixed in a point* is $SO(3)$, the group of proper orthogonal transformations of \mathbb{R}^3 to itself, while the corresponding phase-space is its cotangent bundle, $T^*SO(3)$. The motion of a rigid body is a *geodesic* with respect to a *left-invariant Riemannian metric* (the inertia tensor) on $SO(3)$. The *momentum map* $J : P \rightarrow \mathbb{R}^3$ for the *left* $SO(3)$ -action is *right* translation to the identity. We identify $\mathfrak{so}(3)^*$ with $\mathfrak{so}(3)$ via the *Killing form* and identify \mathbb{R}^3 with $\mathfrak{so}(3)$ via the map $v \mapsto \hat{v}$, where $\hat{v}(w) = v \times w$ (\times being the standard cross product). Points in $\mathfrak{so}(3)^*$ are regarded as the left reduction of $T^*SO(3)$ by $G = SO(3)$ and are the angular momenta as seen from a *body-fixed frame*.

A Segment of a Human-Like Body

A *rigid body with a fixed point* is a basic model of a single segment of the human (or robot) body. This is a left-invariant Hamiltonian mechanical system on the phase-space $T^*SO(3)$. The differentiable structure on $SO(3)$ is defined using the traditional Euler angles $\{\varphi, \psi, \theta\}$. More precisely, a local

chart is

$$(\varphi, \psi, \theta) \in \mathbb{R}^3 \longmapsto A \in SO(3), \quad 0 < \varphi, \psi < 2\pi; \quad 0 < \theta < \pi, \quad \text{where}$$

$$A = \begin{bmatrix} \cos \psi \cos \varphi - \cos \theta \sin \varphi \sin \psi & \cos \psi \cos \varphi + \cos \theta \cos \varphi \sin \psi & \sin \theta \sin \psi \\ -\sin \psi \cos \varphi - \cos \theta \sin \varphi \sin \psi & -\sin \psi \cos \varphi + \cos \theta \cos \varphi \sin \psi & \sin \theta \cos \psi \\ \sin \theta \sin \varphi & -\sin \theta \cos \varphi & \cos \theta \end{bmatrix}$$

The corresponding conjugate momenta are denoted by $p_\varphi, p_\psi, p_\theta$, so $\{\varphi, \psi, \theta, p_\varphi, p_\psi, p_\theta\}$ is the phase-space $T^*SO(3)$. Thus, we have

$$M = T^*SO(3), \quad \omega = dp_\varphi \wedge d\varphi + dp_\psi \wedge d\psi + dp_\theta \wedge d\theta, \quad H = \frac{1}{2}K,$$

$$K = \frac{[(p_\varphi - p_\psi \cos \theta) \sin \psi + p_\theta \sin \theta \cos \psi]^2}{I_1 \sin^2 \theta} + \frac{[(p_\varphi - p_\psi \cos \theta) \cos \psi - p_\theta \sin \theta \sin \psi]^2}{I_2 \sin^2 \theta} + \frac{p_\psi^2}{I_3},$$

where I_1, I_2, I_3 are the moments of inertia, diagonalizing the inertia tensor of the body.

The Hamiltonian equations are

$$\dot{\varphi} = \frac{\partial H}{\partial p_\varphi}, \quad \dot{\psi} = \frac{\partial H}{\partial p_\psi}, \quad \dot{\theta} = \frac{\partial H}{\partial p_\theta},$$

$$\dot{p}_\varphi = -\frac{\partial H}{\partial \varphi}, \quad \dot{p}_\psi = -\frac{\partial H}{\partial \psi}, \quad \dot{p}_\theta = -\frac{\partial H}{\partial \theta}.$$

For each $f, g \in C^\infty(T^*SO(3), \mathbb{R})$ the Poisson bracket is given by

$$\{f, g\}_\omega = \frac{\partial f}{\partial \varphi} \frac{\partial g}{\partial p_\varphi} - \frac{\partial f}{\partial p_\varphi} \frac{\partial g}{\partial \varphi} + \frac{\partial f}{\partial \psi} \frac{\partial g}{\partial p_\psi} - \frac{\partial f}{\partial p_\psi} \frac{\partial g}{\partial \psi}$$

$$+ \frac{\partial f}{\partial \theta} \frac{\partial g}{\partial p_\theta} - \frac{\partial f}{\partial p_\theta} \frac{\partial g}{\partial \theta}.$$

The Heavy Top – Continued

Recall (see (3.8.4.2) above) that the heavy top is by definition a rigid body moving about a fixed point in a 3D space [Putz (1993)]. The rigidity of the top means that the distances between points of the body are fixed

as the body moves. In this case we have

$$\begin{aligned}
 M &= T^*SO(3), \\
 \omega &= dp_\varphi \wedge d\varphi + dp_\psi \wedge d\psi + dp_\theta \wedge d\theta, \\
 H &= \frac{1}{2}K + mgl \cos \theta, \\
 K &= \frac{[(p_\varphi - p_\psi \cos \theta) \sin \psi + p_\theta \sin \theta \cos \psi]^2}{I_1 \sin^2 \theta} \\
 &\quad + \frac{[(p_\varphi - p_\psi \cos \theta) \cos \psi - p_\theta \sin \theta \sin \psi]^2}{I_2 \sin^2 \theta} + \frac{p_\psi^2}{I_3},
 \end{aligned}$$

where I_1, I_2, I_3 are the moments of inertia, m is the total mass, g is the gravitational acceleration and l is the length of the vector determining the center of mass at $t = 0$.

The Hamiltonian equations are

$$\begin{aligned}
 \dot{\varphi} &= \frac{\partial H}{\partial p_\varphi}, & \dot{\psi} &= \frac{\partial H}{\partial p_\psi}, & \dot{\theta} &= \frac{\partial H}{\partial p_\theta}, \\
 \dot{p}_\varphi &= -\frac{\partial H}{\partial \varphi}, & \dot{p}_\psi &= -\frac{\partial H}{\partial \psi}, & \dot{p}_\theta &= -\frac{\partial H}{\partial \theta}.
 \end{aligned}$$

For each $f, g \in C^\infty(T^*SO(3), \mathbb{R})$ the Poisson bracket is given by

$$\begin{aligned}
 \{f, g\}_\omega &= \frac{\partial f}{\partial \varphi} \frac{\partial g}{\partial p_\varphi} - \frac{\partial f}{\partial p_\varphi} \frac{\partial g}{\partial \varphi} + \frac{\partial f}{\partial \psi} \frac{\partial g}{\partial p_\psi} - \frac{\partial f}{\partial p_\psi} \frac{\partial g}{\partial \psi} \\
 &\quad + \frac{\partial f}{\partial \theta} \frac{\partial g}{\partial p_\theta} - \frac{\partial f}{\partial p_\theta} \frac{\partial g}{\partial \theta}.
 \end{aligned}$$

The Hamiltonian H is invariant under rotations about the z -axis, *i.e.*, φ is a cyclic variable, so p_φ is a constant of motion. The momentum map for this S^1 -action is $J(\varphi, \psi, \theta, p_\varphi, p_\psi, p_\theta) = p_\varphi$. The reduced phase-space $J^{-1}(p_\varphi)/S^1$ can be identified with T^*S^2 and it is parameterized by $\{\psi, \theta, p_\psi, p_\theta\}$. The equations of motion for ψ, θ are just Hamiltonian equations for H with p_φ held constant.

Two Coupled Pendula

The configuration space of the system of two coupled pendula in the plane is $T^2 = \{(\theta_1, \theta_2)\}$, where the θ s are the two pendulum angles, the

phase-space is T^*T^2 with its canonical symplectic structure and the Hamiltonian H is given by [Puta (1993)]

$$H = \frac{1}{2}(p_\varphi^2 + p_\psi^2) + V(\sqrt{2}\psi), \quad \text{where}$$

$$\varphi = \frac{\theta_1 + \theta_2}{\sqrt{2}}, \quad \psi = \frac{\theta_1 - \theta_2}{\sqrt{2}}.$$

The circle group S^1 acts on a torus T^2 by $\theta \cdot (\theta_1 + \theta_2) = (\theta + \theta_1, \theta + \theta_2)$ and hence the induced momentum map for the lifted action to T^*T^2 is given by $J(\varphi, \psi, p_\varphi, p_\psi) = p_\varphi$. Therefore, the reduced phase-space $J^{-1}(p_\varphi)/S^1$ is symplectically diffeomorphic to T^*S^1 with its canonical symplectic structure $\omega_\mu = dp_\psi \wedge d\psi$. The reduced Hamiltonian H_μ is $H_\mu = \frac{1}{2}p_\psi^2 + V(\sqrt{2}\psi)$, and Hamiltonian equations for H_μ are

$$\dot{\psi} = p_\psi, \quad \dot{p}_\psi = -\sqrt{2}\dot{V}(\sqrt{2}\psi).$$

The Plane 2-Body Problem

The plane two body problem can be formulated as the triple (M, ω, H) where [Puta (1993)]

$$M = T^*((0, \infty) \times S^1), \quad \omega = dp_r \wedge dr + dp_\theta \wedge d\theta,$$

$$H = (p_r^2 + p_\theta^2)/r^2 - 1/r.$$

The Lie group $G = SO(2) \simeq S^1$ acts on the configuration space $M = (0, \infty) \times S^1$ by rotations, i.e., if $R_\varphi \in SO(2)$ then

$$\phi : (R_\varphi, (r, \theta)) \mapsto (r, \theta + \varphi, p_r, p_\theta).$$

The corresponding momentum map is

$$J(r, \theta, p_r, p_\theta) = p_\theta.$$

The 3-Body Problem

There is a vast literature on the *restricted three-body problem* (see [Meyer and Hall (1992)]). Among other things, there are investigations of the equilibria points and their stability, investigations of the existence, stability and bifurcation of periodic orbits, and investigations of collisions and ejection orbits. The restricted problem is said to be a limit of the *full three-body problem* as one of the masses tends to zero, and so to each result

for the restricted problem there should be a corresponding result for the full three-body problem.

The restricted three-body problem is a Hamiltonian system of differential equations which describes the motion of an infinitesimal particle (the *satellite*) moving under the gravitational influence of two particles of finite mass (the *primaries*) which are moving on a circular orbit of the Kepler problem [Meyer (2005)].

Since the motion of the primaries is given, the restricted problem has two DOF for the planar problem and three DOF for the spatial problem. However, the full problem has six DOF in the planar case and nine DOF in the spatial case. Thus, at first the restricted problem seems too small to reflect the full complexity of the full problem; but when the symmetries of the full problem are taken into account the dimension gap narrows considerably.

The Hamiltonian of the full problem is invariant under Euclidean motions, i.e., translations and rotations, which begets the integrals of linear and angular momentum. Translations and rotations induce ignorable coordinates. Holding the integrals fixed and dropping the ignorable coordinates reduces the full problem from six to three DOF in the planar case and from nine to four DOF in the spatial case. Thus the full problem on the reduced space is only one DOF larger than the restricted problem in either the planar or the spatial case [Meyer (2005)].

The full 3-body problem in 3D space has 9 DOF. By placing the center of mass at the origin and setting linear momentum equal to zero the problem reduces one with six DOF. This can be done using Jacobi coordinates. The Hamiltonian of the full 3-body problem in rotating (about the z -axis) Jacobi coordinates $(u_0, u_1, u_2, v_0, v_1, v_2)$ is

$$H = \frac{\|v_0\|^2}{2M_0} - u_0^T J v_0 + \frac{\|v_1\|^2}{2M_1} - u_1^T J v_1 - \frac{m_0 m_1}{\|u_1\|} \\ + \frac{\|v_2\|^2}{2M_2} - u_2^T J v_2 - \frac{m_1 m_2}{\|u_2 - \alpha_0 u_1\|} - \frac{m_2 m_0}{\|u_2 + \alpha_1 u_1\|}$$

where $u_i, v_i \in \mathbb{R}^3$,

$$M_0 = m_0 + m_1 + m_2, \quad M_1 = m_0 m_1 / (m_0 + m_1), \\ M_2 = m_2 (m_0 + m_1) / (m_0 + m_1 + m_2), \\ \alpha_0 = m_0 / (m_0 + m_1), \quad \alpha_1 = m_1 / (m_0 + m_1),$$

$$\text{and } J = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

In these coordinates u_0 is the center of mass, v_0 is total linear momentum, and total angular momentum is: $A = u_0 \times v_0 + u_1 \times v_1 + u_2 \times v_2$. See [Meyer and Hall (1992)] for further details.

n -DOF Hamiltonian Dynamics

Classically, n -DOF Hamiltonian dynamics combines the ideas of differential equations and variational principles (see [Abraham and Marsden (1978); Arnold (1989); Marsden and Ratiu (1999); Wiggins (1990)]). As Hamiltonian first realized, many of the systems of mechanics and optics can be put into the special form (compare (3.164))

$$\dot{q}^i = \frac{\partial H}{\partial p_i}(q^i, p_i, t), \quad \dot{p}_i = -\frac{\partial H}{\partial q^i}(q^i, p_i, t), \quad (i = 1, \dots, n),$$

or an associated variational form (summing upon the repeated index is used in the following text)

$$\delta \int (p_i dq^i - H) dt = 0.$$

Here the state of the system is given as a point $(q^1, \dots, q^n, p_1, \dots, p_n)$ in *phase-space*, the q 's are the configuration coordinates, the p 's are the momenta, t is time, and $H = H(q^i, p_i, t)$ is a total-energy function called Hamiltonian. The variables (q^i, p_i) are called *canonical* coordinates.

If $H = H(q^i, p_i)$ does not depend explicitly on time, the system is said to be *autonomous*. In this case, it is easy to verify that H is conserved. The search for other conserved quantities led to a new notion of solving Hamiltonian systems. Instead of finding formulae for the coordinates as a function of time, one searches for constants of the motion (*integrals*). If one can find n integrals $I_i(q^i, p_i)$ which are in *involution*:

$$[I_i, I_j] = \frac{\partial I_i}{\partial q^k} \frac{\partial I_j}{\partial p_k} - \frac{\partial I_i}{\partial p_k} \frac{\partial I_j}{\partial q^k} = 0, \quad (i \neq j),$$

and *independent* (the vectors ∇I_i are independent 'almost everywhere'), then associated variables ϕ_i can be derived which evolve linearly in time: $\dot{\phi}^i = \frac{\partial H}{\partial I_i}(I_i)$.

Such a system is *integrable* in the sense of Liouville [Arnold (1989)]. If the sets $I = \text{const}$ are bounded, then they are n D *tori* T^n in phase-space.

Choosing irreducible cycles, γ_i , on the tori, one can define a preferred set of integrals $J_i = \int_{\gamma_i} p_i dq^i$, called *action variables*, for which the corresponding ϕ_i are *angle variables* mod 1 on T^n . The quantities $\omega^i(J) = \frac{\partial H}{\partial J_i}(J_i)$ are called the *frequencies* on T^n .

Another feature of Hamiltonian systems noticed by Liouville is the preservation of phase-space volume $\int (dq)^n (dp)^n$. A more general result is that Poincaré's integral $\int p_i dq^i$ is conserved around any loop following the flow [Arnold (1989)]. This is the property that really distinguishes Hamiltonian differential equations from general ones.

The major problem with the notion of integrability is that *most systems are not integrable*. This was first appreciated when Poincaré proved that the *circular restricted three-body problem has no integral analytic in the mass ratio*. The perturbation expansions which gave excellent predictions of motion of the planets do not converge. The basic reason is that among the invariant tori of integrable systems is a dense subset on which the frequencies ω^i are *commensurate*, i.e., $m_i \omega^i = 0$ for some non-zero integer vector m_i . However, most systems have no commensurate tori, because they can be destroyed by arbitrarily small perturbation.

Poincaré went on to examine what really does happen. The key technique he used was geometrical analysis: instead of manipulating formulae for canonical transformations as Jacobi and others did, he pictured the *orbits* in phase-space. An important step in this *qualitative ODE theory* was the idea of *surface of section*. If Σ is a *codimension-one* surface (i.e., of dimension one less than that of the phase-space) transverse to a *flow*, then the sequence $\{x_j\}$ of successive intersections of an orbit with Σ gives a lot of information about that orbit. For example, if $\{x_j\}$ is periodic then it corresponds to a periodic orbit. If $\{x_j\}$ is confined to a subset of codimension m on Σ then so is the orbit of the flow, etc.. The flow induces a mapping of Σ to itself; the map takes a point in Σ to the point at which it first returns to Σ (assuming there is one). Since the surface of section has one dimension less than the phase-space it is easier to picture the dynamics of the return map than the flow. In fact, for Hamiltonian systems one can do even better; since H is conserved, Σ decomposes into a one-parameter family of codimension two surfaces parameterized by the value of the energy, a reduction of two dimensions.

This led Poincaré to the ideas of *stable* and *unstable manifolds* for *hyperbolic* periodic orbits, which are extensions of the stable and unstable *eigenspaces* for associated linear systems, and their intersections, known as

hetero- and homo-clinic points, whose orbits converge to one periodic orbit in the past and to another (or the same) in the future. He showed that having intersected once, the invariant manifolds must intersect infinitely often. Moreover the existence of one heteroclinic orbit implies the existence of an infinity of others.

The distance between the stable and unstable manifolds can be quantified by Melnikov's integral. This leads to a technique for proving the non-existence of integrals for a slightly perturbed, integrable Hamiltonian.

For integrable systems, nearby orbits separate linearly in time. However, dynamical systems can have exponentially separating orbits. Let δx be a tangent vector at the phase-space point x and δx_t be the evolved vector following the orbit of x . Then, recall that the average rate of exponentiation of δx_t is the *Lyapunov exponent* λ (see, e.g., [Chen and Dong (1998)])

$$\lambda(x, \delta x) = \lim_{t \rightarrow \infty} 1/t \ln |\delta x_t|.$$

If λ is nonzero, then the predictions one can make will be valid for a time only logarithmic in the precision. Therefore, although deterministic in principle, a system need not be predictable in practice.

A concrete example of the complexity of behavior of typical Hamiltonian systems is provided by the 'horseshoe', a type of invariant set found near homoclinic orbits. Its points can be labelled by doubly infinite sequences of 0's and 1's corresponding to which half of a horseshoe shaped set the orbit is in at successive times. For every sequence, no matter how complicated, there is an orbit which has that symbol sequence. This implies, e.g., that a simple pendulum in a sufficiently strongly modulated time-periodic gravitational field has an initial condition such that the pendulum will turn over once each period when there is 1 in the sequence and not if there is a 0 for any sequence of 0's and 1's.

3.12.3.3 *Hamilton-Poisson Mechanics*

Now, instead of using symplectic structures arising in Hamiltonian mechanics, we propose the more general *Poisson manifold* $(\mathfrak{g}^*, \{F, G\})$. Here \mathfrak{g}^* is a chosen Lie algebra with a (\pm) *Lie-Poisson bracket* $\{F, G\}_{\pm}(\mu)$ and carries an abstract *Poisson evolution equation* $\dot{F} = \{F, H\}$. This approach is well-defined in both the finite- and the infinite-dimensional case. It is equivalent to the strong symplectic approach when this exists and offers a viable formulation for Poisson manifolds which are not symplectic (for technical details, see [Weinstein (1990); Abraham *et al.* (1988);

Marsden and Ratiu (1999); Puta (1993); Ivancevic and Pearce (2001a)]).

Let E_1 and E_2 be Banach spaces. A continuous bilinear functional $\langle, \rangle: E_1 \times E_2 \rightarrow \mathbb{R}$ is nondegenerate if $\langle x, y \rangle = 0$ implies $x = 0$ and $y = 0$ for all $x \in E_1$ and $y \in E_2$. We say E_1 and E_2 are in *duality* if there is a nondegenerate bilinear functional $\langle, \rangle: E_1 \times E_2 \rightarrow \mathbb{R}$. This functional is also referred to as an L^2 -pairing of E_1 with E_2 .

Recall that a *Lie algebra* consists of a vector space \mathfrak{g} (usually a Banach space) carrying a bilinear skew-symmetric operation $[\cdot, \cdot]: \mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g}$, called the *commutator* or Lie bracket. This represents a pairing $[\xi, \eta] = \xi\eta - \eta\xi$ of elements $\xi, \eta \in \mathfrak{g}$ and satisfies *Jacobi identity*

$$[[\xi, \eta], \mu] + [[\eta, \mu], \xi] + [[\mu, \xi], \eta] = 0.$$

Let \mathfrak{g} be a (finite- or infinite-dimensional) Lie algebra and \mathfrak{g}^* its dual Lie algebra, that is, the vector space L^2 paired with \mathfrak{g} via the inner product $\langle, \rangle: \mathfrak{g}^* \times \mathfrak{g} \rightarrow \mathbb{R}$. If \mathfrak{g} is nD , this pairing reduces to the usual action (interior product) of forms on vectors. The standard way of describing any nD Lie algebra \mathfrak{g} is to give its n^3 *Lie structural constants* γ_{ij}^k , defined by $[\xi_i, \xi_j] = \gamma_{ij}^k \xi_k$, in some basis ξ_i , ($i = 1, \dots, n$)

For any two smooth functions $F: \mathfrak{g}^* \rightarrow \mathbb{R}$, we define the *Fréchet derivative* D on the space $L(\mathfrak{g}^*, \mathbb{R})$ of all linear diffeomorphisms from \mathfrak{g}^* to \mathbb{R} as a map $DF: \mathfrak{g}^* \rightarrow L(\mathfrak{g}^*, \mathbb{R})$; $\mu \mapsto DF(\mu)$. Further, we define the *functional derivative* $\delta F/\delta\mu \in \mathfrak{g}$ by

$$DF(\mu) \cdot \delta\mu = \langle \delta\mu, \frac{\delta F}{\delta\mu} \rangle$$

with arbitrary ‘variations’ $\delta\mu \in \mathfrak{g}^*$.

For any two smooth functions $F, G: \mathfrak{g}^* \rightarrow \mathbb{R}$, we define the (\pm) *Lie-Poisson bracket* by

$$\{F, G\}_{\pm}(\mu) = \pm \langle \mu, \left[\frac{\delta F}{\delta\mu}, \frac{\delta G}{\delta\mu} \right] \rangle. \quad (3.1)$$

Here $\mu \in \mathfrak{g}^*$, $[\xi, \mu]$ is the Lie bracket in \mathfrak{g} and $\delta F/\delta\mu, \delta G/\delta\mu \in \mathfrak{g}$ are the functional derivatives of F and G .

The (\pm) Lie-Poisson bracket (3.1) is clearly a bilinear and skew-symmetric operation. It also satisfies the Jacobi identity

$$\{\{F, G\}, H\}_{\pm}(\mu) + \{\{G, H\}, F\}_{\pm}(\mu) + \{\{H, F\}, G\}_{\pm}(\mu) = 0$$

thus confirming that \mathfrak{g}^* is a Lie algebra, as well as Leibniz' rule

$$\{FG, H\}_{\pm}(\mu) = F\{G, H\}_{\pm}(\mu) + G\{F, H\}_{\pm}(\mu). \quad (3.177)$$

If \mathfrak{g} is a nD phase-space manifold with structure constants γ_{ij}^k , the (\pm) Lie–Poisson bracket (3.177) becomes

$$\{F, G\}_{\pm}(\mu) = \pm \mu_k \gamma_{ij}^k \frac{\delta F}{\delta \mu_i} \frac{\delta G}{\delta \mu_j}. \quad (3.178)$$

The (\pm) Lie–Poisson bracket represents a Lie-algebra generalization of the classical nD Poisson bracket $[F, G] = \omega(X_f, X_g)$ on the symplectic phase-space manifold (P, ω) for any two real-valued smooth functions $F, G : P \rightarrow \mathbb{R}$.

As in the classical case, any two smooth functions $F, G : \mathfrak{g}^* \rightarrow \mathbb{R}$ are *in involution* if $\{F, G\}_{\pm}(\mu) = 0$.

The Lie–Poisson Theorem states that a Lie algebra \mathfrak{g}^* with a \pm Lie–Poisson bracket $\{F, G\}_{\pm}(\mu)$ represents a Poisson manifold $(\mathfrak{g}^*, \{F, G\}_{\pm}(\mu))$.

Given a smooth Hamiltonian function $H : \mathfrak{g}^* \rightarrow \mathbb{R}$ on the Poisson manifold $(\mathfrak{g}^*, \{F, G\}_{\pm}(\mu))$, the time evolution of any smooth function $F : \mathfrak{g}^* \rightarrow \mathbb{R}$ is given by the abstract *Poisson evolution equation*

$$\dot{F} = \{F, H\}. \quad (3.179)$$

3.12.3.4 Completely Integrable Hamiltonian Systems

In order to integrate a system of $2n$ ODEs, we must know $2n$ first integrals. It turns out that if we are given a canonical system of ODEs, it is often sufficient to know only n first integrals [Arnold (1989)].

Liouville Theorem on Completely Integrable Systems

Recall that a function F is a first integral of a system Ξ with Hamiltonian function H iff H and F are in involution on the system's phase-space P (which is the cotangent bundle of the system's configuration manifold T^*M), i.e., iff the Poisson bracket of H and F is identically equal to zero on P , $\{H, F\} \equiv 0$.

Liouville proved that if, in a system Ξ with n DOF (i.e., with a $2nD$ phase-space $P = T^*M$), n independent first integrals in involution are known, then the system is integrable by quadratures.

Here is the exact formulation of the *Liouville Theorem* [Arnold (1989)]: Suppose that we are given n functions in involution on a symplectic $2nD$

manifold:

$$F_1, \dots, F_n; \quad \{F_i, F_j\} \equiv 0, \quad (i, j = 1, \dots, n).$$

Consider a level set of the functions F_i :

$$M_f = \{x : F_i(x) = f_i\}, \quad (i = 1, \dots, n).$$

Assume that the n functions F_i are independent on M_f (i.e., the n 1-forms dF_i are linearly independent at each point of M_f). Then

1. M_f is a smooth manifold, invariant under the phase-flow with Hamiltonian function $H = F_1$.

2. If the manifold M_f is compact and connected, then it is diffeomorphic to the n -torus

$$T^n = \{(\varphi^1, \dots, \varphi^n) \bmod 2\pi\}.$$

3. The phase-flow with Hamiltonian function H determines a conditionally periodic motion on M_f , i.e., in angular coordinates $\varphi^i = (\varphi^1, \dots, \varphi^n)$ we have

$$\dot{\varphi}^i = \omega^i, \quad \omega^i = \omega^i(f_i), \quad (i = 1, \dots, n).$$

4. The canonical equations with Hamiltonian function H can be integrated by quadratures.

For the proof of this Theorem see [Arnold (1989)].

As an example with 3 DOF, we consider a heavy symmetrical Lagrangian top fixed at a point on its axis. Three first integrals are immediately obvious: H , M_z and M_3 . It is easy to verify that the integrals M_z and M_3 are in involution. Furthermore, the manifold $H = h$ in the phase-space is compact. Therefore, we can say without any calculations that the motion of the top is conditionally periodic: the phase trajectories fill up the 3D torus T^3 , given by: $H = c_1$, $M_z = c_2$, $M_3 = c_3$. The corresponding three frequencies are called frequencies of fundamental rotation, precession, and nutation.

Other examples arise from the following observation: *if a canonical system can be integrated by the method of Hamiltonian-Jacobi, then it has n first integrals in involution.* The method consists of a canonical transformation $(p_i, q^i) \rightarrow (P_i, Q^i)$ such that the Q^i are first integrals, while the functions Q^i and Q^i are in involution.

The Liouville Theorem, as formulated above, covers all the problems of dynamics which have been integrated to the present day [Arnold (1989)].

Action–Angle Variables

Under the hypothesis of the Liouville Theorem, we can find symplectic coordinates (I_i, φ^i) such that the first integrals F_i depend only on I_i and φ^i (for $i = 1, \dots, n$) are angular coordinates on the n -torus $T^n \simeq M_f = \{x : F_i(x) = f_i\}$, which is invariant with respect to the phase-flow. We choose angular coordinates φ^i on M_f so that the phase-flow with Hamiltonian function $H = F_1$ takes an especially simple form [Arnold (1989)]:

$$\dot{\varphi}^i = \omega^i(f_i), \quad \varphi^i(t) = \varphi^i(0) + \omega^i t.$$

Now we look at a neighborhood of the n -manifold $M_f = T^n$ in the system's $2n$ D phase-space P .

In the coordinates (F_i, φ^i) the phase-flow with Hamiltonian function $H = F_1$ can be written in the form of the simple system of $2n$ ODEs

$$\dot{F}_i = 0, \quad \dot{\varphi}^i = \omega^i(F_i), \quad (i = 1, \dots, n), \quad (3.180)$$

which is easily integrated: $F_i(t) = F_i(0)$, $\varphi^i(t) = \varphi^i(0) + \omega^i(F_i(0))t$.

Thus, in order to integrate explicitly the original canonical system of ODEs, it is sufficient to find the variables φ^i in explicit form. It turns out that this can be done using only quadratures. A construction of the variables φ^i is given below [Arnold (1989)].

In general, the variables (F_i, φ^i) are not symplectic coordinates. However, there are functions of F_i , which we denote by $I_i = I_i(F_i)$, ($i = 1, \dots, n$), such that the variables (I_i, φ^i) are symplectic coordinates: the original symplectic structure $dp_i \wedge dq^i$ is expressed in them as $dI_i \wedge d\varphi^i$. The variables I_i have physical dimensions of action and are called action variables; together with the angle variables φ^i they form the *action–angle system of canonical coordinates* in a neighborhood of $M_f = T^n$.

The quantities I_i are first integrals of the system with Hamiltonian function $H = F_1$, since they are functions of the first integrals F_i . In turn, the variables F_i can be expressed in terms of I_i and, in particular, $H = F_1 = H(I_i)$. In action–angle variables, the ODEs of our flow (3.180) have the form

$$\dot{I}_i = 0, \quad \dot{\varphi}^i = \omega^i(I_i), \quad (i = 1, \dots, n).$$

A system with one DOF in the phase plane (p, q) is given by the Hamiltonian function $H(p, q)$. In order to construct the *action–angle variables*, we look for a canonical transformation $(p, q) \rightarrow (I, \varphi)$ satisfying the two

conditions:

$$I = I(h), \quad \oint_{M_h} d\varphi = 2\pi. \quad (3.181)$$

The *action variable* in the system with one DOF given by the Hamiltonian function $H(p, q)$ is the quantity

$$I(h) = \frac{1}{2\pi} \Pi(h) = \frac{1}{2\pi} \oint_{M_h} p dq,$$

which is the area bounded by the phase curve $H = h$. Arnold states the following Theorem: Set $S(I, q) = \int_{q_0}^q p dq|_{H=h(I)}$ is a *generating function*. Then a canonical transformation $(p, q) \rightarrow (I, \varphi)$ satisfying conditions (3.181) is given by

$$p = \frac{\partial S(I, q)}{\partial q}, \quad \varphi = \frac{\partial S(I, q)}{\partial I}, \quad H\left(\frac{\partial S(I, q)}{\partial q}, q\right) = h(I).$$

We turn now to systems with n DOF given in $\mathbb{R}^{2n} = \{(p_i, q^i), i = 1, \dots, n\}$ by a Hamiltonian function $H(p_i, q^i)$ and having n first integrals in involution $F_1 = H, F_2, \dots, F_n$. Let $\gamma_1, \dots, \gamma_n$ be a basis for the 1D cycles on the torus $M_f = T^n$ (the increase of the coordinate φ^i on the cycle γ_j is equal to 2π if $i = j$ and 0 if $i \neq j$). We set

$$I_i(f_i) = \frac{1}{2\pi} \oint_{M_h} p_i dq^i, \quad (i = 1, \dots, n). \quad (3.182)$$

The n quantities $I_i(f_i)$ given by formula (3.182) are called the *action variables* [Arnold (1989)].

We assume now that, for the given values f_i of the n integrals F_i , the n quantities I_i are independent, $\det(\partial I_i / \partial f_i)|_{f_i} \neq 0$. Then in a neighborhood of the torus $M_f = T^n$ we can take the variables I_i, φ^i as symplectic coordinates, i.e., the transformation $(p_i, q^i) \rightarrow (I_i, \varphi^i)$ is canonical, i.e.,

$$dp_i \wedge dq^i = dI_i \wedge d\varphi^i, \quad (i = 1, \dots, n).$$

Now, let m be a point on M_f , in a neighborhood of which the n variables q^i are coordinates of M_f , such that the submanifold $M_f \subset \mathbb{R}^{2n}$ is given by n equations of the form $p_i = p_i(I_i, q^i)$, $q^i(m) = q_0^i$. In a simply-connected neighborhood of the point q_0^i a single-valued function is defined,

$$S(I_i, q^i) = \int_{q_0}^q p_i(I_i, q^i) dq^i,$$

and we can use it as the generating function of a canonical transformation $(p_i, q^i) \rightarrow (I_i, \varphi^i)$:

$$p_i = \frac{\partial S}{\partial q^i}, \quad \varphi^i = \frac{\partial S}{\partial I_i}.$$

A Universal Model for Completely Integrable Systems

A Hamiltonian system on a $2n$ D symplectic manifold M is said to be *completely integrable* if it has n first integrals in involution, which are functionally independent on some open dense submanifold of M . This definition of a completely integrable system is usually found, with some minor variants, in any modern text on symplectic mechanics [Arnold (1989); Abraham and Marsden (1978); Libermann and Marle (1987); Marmo *et. al.* (1995); Thirring (1979)].

Starting with this definition, one uses the so-called *Liouville–Arnold Theorem* to introduce action–angle variables and write the Hamiltonian system in the form

$$\dot{I}^k = 0, \quad \dot{\phi}_k = \frac{\partial H}{\partial I_k} = \nu_k(I),$$

where $k \in \{1, \dots, n\}$. The corresponding flow is given by

$$I^k(t) = I^k(0), \quad \phi_k(t) = \phi_k(0) + \nu_k t. \quad (3.183)$$

The main interest in completely integrable systems relies on the fact that they can be integrated by quadratures [Arnold (1989)].

However, it is clear that even if $\nu_k dI^k$ is not an exact (or even a closed) 1-form, as long as $\dot{\nu}_k = 0$, the system can always be integrated by quadratures.

If we consider the Abelian Lie group \mathbb{R}^n , we can construct a Hamiltonian action of \mathbb{R}^n on $T^*\mathbb{R}^n$ induced by the group addition: $\mathbb{R}^n \times T^*\mathbb{R}^n \rightarrow T^*\mathbb{R}^n$. This can be generalized to the Hamiltonian action [Alekseevsky *et. al.* (1997)]

$$\mathbb{R}^n \times T^*(\mathbb{R}^k \times T^{n-k}) \rightarrow T^*(\mathbb{R}^k \times T^{n-k}),$$

of \mathbb{R}^n , where T^m stands for the m D torus, and reduces to $\mathbb{R}^n \times T^*T^n$ or $T^n \times T^*T^n$, when $k = 0$.

By using the standard symplectic structure on $T^*\mathbb{R}^n$, we find the momentum map $\mu : T^*\mathbb{R}^n \rightarrow (\mathbb{R}^n)^s$, $(q, p) \mapsto p$, induced by the natural action of \mathbb{R}^n on itself via translations, which is a Poisson map if $(\mathbb{R}^n)^s$ is

with the (trivial) natural Poisson structure of the dual of a Lie algebra. It is now clear that any function on $(\mathbb{R}^n)^s$, when pulled back to $T^*\mathbb{R}^n$ or T^*T^n , induces a Hamiltonian system which is completely integrable (in the Liouville sense). Because the level sets of this function carry on the action of \mathbb{R}^n , the completely integrable system induces a 1D subgroup of the action of \mathbb{R}^n on the given level set. However, the specific subgroup will depend on the particular level set, i.e., the ‘frequencies’ are first integrals. The property of being integrable by quadratures is captured by the fact that it is a subgroup of the \mathbb{R}^n -action on each level set.

It is now clear, how we can preserve this property, while giving up the requirement that our system is Hamiltonian. We can indeed consider any 1-form η on $(\mathbb{R}^n)^s$ and pull it back to $T^*\mathbb{R}^n$ or T^*T^n , then associated vector-field $\Gamma_\eta = \Lambda_0(\mu^s(\eta))$, where Λ_0 is the canonical Poisson structure in the cotangent bundle, is no more Hamiltonian, but it is still integrable by quadratures. In action-angle variables, if $\eta = \nu_k dI^k$ is the 1-form on $(\mathbb{R}^n)^s$, the associated equations of motion on T^*T^n will be [Alekseevsky *et. al.* (1997)]

$$\dot{I}^k = 0, \quad \dot{\phi}_k = \nu_k,$$

with $\dot{\nu}_k = 0$, therefore the flow will be as in (3.183), even though $\partial_{I^j}\nu_k \neq \partial_{I^k}\nu_j$.

We can now generalize this construction to any Lie group G . We consider the Hamiltonian action $G \times T^*G \rightarrow T^*G$, of G on the cotangent bundle, induced by the right action of G on itself. The associated momentum map $\mu : T^*G \simeq \mathcal{G}^s \times G \longrightarrow \mathcal{G}^s$. It is a Poisson map with respect to the natural Poisson structure on \mathcal{G}^s (see, e.g., [Alekseevsky *et. al.* (1994); Libermann and Marle (1987)]).

Now, we consider any differential 1-form η on \mathcal{G}^s which is annihilated by the natural Poisson structure $\Lambda_{\mathcal{G}^s}$ on \mathcal{G}^s associated with the Lie bracket. Such form we call a *Casimir form*. We define the vector-field $\Gamma_\eta = \Lambda_0(\mu^s(\eta))$. Then, the corresponding dynamical system can be written as [Alekseevsky *et. al.* (1997)]

$$g^{-1}\dot{g} = \eta(g, p) = \eta(p), \quad \dot{p} = 0,$$

since $\omega_0 = d(\langle p, g^{-1}dg \rangle)$ (see [Alekseevsky *et. al.* (1994)]). Here we interpret the covector $\eta(p)$ on \mathcal{G}^s as a vector of \mathcal{G} . Again, our system can be integrated by quadratures, because on each level set, get by fixing p 's in \mathcal{G}^s , our dynamical system coincides with a one-parameter group of the

action of G on that particular level set.

We give a familiar example: the rigid rotator and its generalizations [Alekseevsky *et. al.* (1997)]. In the case of $G = \mathrm{SO}(3)$ the (right) momentum map

$$\mu : T^*\mathrm{SO}(3) \longrightarrow \mathfrak{so}(3)^*$$

is a Poisson map onto $\mathfrak{so}(3)^*$ with the linear Poisson structure

$$\Lambda_{\mathfrak{so}(3)^*} = \varepsilon^{ijk} p_i \partial_{p_j} \otimes \partial_{p_k}.$$

Casimir 1-forms for $\Lambda_{\mathrm{so}(3)^*}$ read $\eta = FdH_0$, where $H_0 = \sum p_i^2/2$ is the ‘free Hamiltonian’ and $F = F(p)$ is an arbitrary function. Clearly, FdH_0 is not a closed form in general, but (p_i) are first integrals for the dynamical system $\Gamma_\eta = \Lambda_0(\mu^s(\eta))$. It is easy to see that

$$\Gamma_\eta = F(p)\Gamma_0 = F(p)p_i \widehat{X}_i,$$

where \widehat{X}_i are left-invariant vector-fields on $\mathrm{SO}(3)$, corresponding to the basis (X_i) of $\mathfrak{so}(3)$ identified with (dp_i) . Here we used the identification $T^*\mathrm{SO}(3) \simeq \mathrm{SO}(3) \times \mathfrak{so}(3)^*$ given by the momentum map μ . In other words, the dynamics is given by

$$\dot{p}_i = 0, \quad g^{-1}\dot{g} = F(p)p_i X_i \in \mathfrak{so}(3),$$

and it is completely integrable, since it reduces to left-invariant dynamics on $\mathrm{SO}(3)$ for every value of p . We recognize the usual isotropic rigid rotator, when $F(p) = 1$.

We can generalize our construction once more, replacing the cotangent bundle T^*G by its deformation, namely a group double $D(G, \Lambda_G)$ associated with a Lie–Poisson structure Λ_G on G (see e.g., [Lu (1990)]). This double, denoted simply by D , carry on a natural *Poisson tensor-field* Λ_D^+ which is non-degenerate on the open-dense subset $D^+ = G \cdot G^s \cap G^s \cdot G$ of D (here $G^s \subset D$ is the dual group of G with respect to Λ_G). We refer to D as being *complete* if $D^+ = D$. Identifying D with $G \times G^s$ if D is complete (or D^+ with an open submanifold of $G \times G^s$ in general case; we assume completeness for simplicity) via the group product, we can write Λ_D^+ in ‘coordinates’ $(g, u) \in G \times G^s$ in the form [Alekseevsky *et. al.* (1997)]

$$\Lambda_D^+(g, u) = \Lambda_G(g) + \Lambda_{G^s}(u) - X_i^l(g) \wedge Y_i^r(u), \quad (3.184)$$

where X_i^l and Y_i^r are, respectively, the left- and right-invariant vector-fields on G and G^s relative to dual bases X_i and Y_i in the Lie algebras \mathcal{G}

and \mathcal{G}^s , and where Λ_G and Λ_{G^*} are the corresponding Lie–Poisson tensors on G and G^s (see [Lu (1990)]). It is clear now that the projections μ_{G^*} and μ_G of (D, Λ_D^+) onto (G, Λ_G) and (G^s, Λ_{G^*}) , respectively, are Poisson maps. Note that we get the cotangent bundle $(D, \Lambda_D^+) = (T^*G, \Lambda_0)$ if we put $\Lambda_G = 0$.

The group G acts on (D, Λ_D^+) by left translations which, in general are not canonical transformations. However, this is a Poisson action with respect to the inner Poisson structure Λ_G on G , which is sufficient to develop the momentum map reduction theory (see [Lu (1991)]). For our purposes, let us take a Casimir 1-form η for Λ_{G^*} , i.e., $\Lambda_{G^*}(\eta) = 0$. By means of the momentum map

$\mu_{G^*} : D \longrightarrow G^s$, we define the vector-field on D [Alekseevsky *et. al.* (1997)]:

$$\Gamma_\eta = \Lambda_D^+(\mu_{G^*}^s(\eta)).$$

In ‘coordinates’ (g, u) , due to the fact that η is a Casimir, we get

$$\Gamma_\eta(g, u) = \langle Y_i^r, \eta \rangle (u) X_i^l(g),$$

so that Γ_η is associated with the *Legendre map*

$$L_\eta : D \simeq G \times G^s \longrightarrow TG \simeq G \times \mathcal{G}, \quad L_\eta(g, u) = \langle Y_i^r, \eta \rangle (u) X_i,$$

which can be viewed also as a map $L_\eta : G^s \longrightarrow \mathcal{G}$. Thus we get the following Theorem [Alekseevsky *et. al.* (1997)]: The dynamics Γ_η on the group double $D(G, \Lambda_G)$, associated with a 1-form η which is a Casimir for the Lie–Poisson structure Λ_{G^*} on the dual group, is given by the system of equations

$$\dot{u} = 0, \quad g^{-1}\dot{g} = \langle Y_i^r, \eta \rangle (u) X_i \in \mathcal{G},$$

and is therefore completely integrable by quadratures.

We have seen that if we concentrate on the possibility of integrating our system by quadratures, then we can do without the requirement that the system is Hamiltonian.

By considering again the equations of motion in *action–angle variables*, we classically have

$$\dot{I}^k = 0, \quad \dot{\phi}_k = \nu^k(I).$$

Clearly, if we have

$$\dot{I}^k = F_k(I), \quad \dot{\phi}_k = A_k^j(I)\phi_j, \quad (3.185)$$

and we are able to integrate the first equation by quadratures, we again have the possibility to integrate by quadratures the system (3.185), if only the matrices $(A_k^j(I(t)))$ commute [Alekseevsky *et. al.* (1997)]:

$$\phi(t) = \exp \left(\int_0^t A(I(s)) ds \right) \phi_0.$$

Because ϕ_k are discontinues functions on the torus, we have to be more careful here. However, we show how this idea works for double groups. In the case when the 1-form η on G^s is not a Casimir 1-form for the Lie-Poisson structure Λ_{G^*} , we get, in view of (3.184),

$$\Gamma_\eta(g, u) = \langle Y_i^r, \eta \rangle (u) X_i^l(g) + \Lambda_{G^*}(\eta)(u).$$

Now, the momenta evolve according to the dynamics $\Lambda_{G^*}(\eta)$ on G^s (which can be interpreted, as we will see later, as being associated with an interaction of the system with an external field) and ‘control’ the evolution of the field of velocities on G (being left-invariant for a fixed time) by a ‘variation of constants’. Let us summarize our observations in the following Theorem [Alekseevsky *et. al.* (1997)]: The vector-field Γ_η on the double group $D(G, \Lambda_G)$, associated with a 1-form η on G^s , defines the following dynamics

$$\dot{u} = \Lambda_{G^*}(\eta)(u), \quad g^{-1}\dot{g} = \langle Y_i^r, \alpha \rangle (u) X_i \in \mathcal{G}, \quad (3.186)$$

and is therefore completely integrable, if only we are able to integrate the equation (3.186) and $\langle Y_i^r, \eta \rangle (u(t)) X_i$ lie in a commutative subalgebra of \mathcal{G} for all t .

Finally, we can weaken the assumptions of the previous Theorem. It is sufficient to assume [Alekseevsky *et. al.* (1997)] that

$$g^{-1}\dot{g}(t) = \exp(tX)A(t)\exp(-tX),$$

for some $A(t)$, $X \in \mathcal{G}$, such that $X + A(t)$ lie in a commutative subalgebra of \mathcal{G} for all t (e.g., $A(t) = \text{const}$), to assure that (3.186) is integrable by quadratures. Indeed, in the new variable

$$g_1(t) = \exp(-tX)g(t)\exp(tX),$$

the equation (3.186) reads

$$\dot{g}_1(t) = g_1(t)(X + A(t)) - Xg_1(t),$$

and, since the right- and the left-multiplications commute, we find that [Alekseevsky *et. al.* (1997)]

$$g(t) = g_0 \exp \left(tX + \int_0^t A(s) ds \right) \exp(-tX).$$

This procedure is similar to what is known as the Dirac interaction picture in the quantum evolution.

3.12.3.5 Momentum Map and Symplectic Reduction

Let (M, ω) be a connected symplectic manifold and $\phi : G \times M \rightarrow M$ a symplectic action of the Lie group G on M , that is, for each $g \in G$ the map $\phi_g : M \rightarrow M$ is a symplectic diffeomorphism. If for each $\xi \in \mathfrak{g}$ there exists a globally defined function $\hat{J}(\xi) : M \rightarrow \mathbb{R}$ such that $\xi_M = X_{\hat{J}(\xi)}$, then the map $J : M \rightarrow \mathfrak{g}^*$, given by

$$J : x \in M \mapsto J(x) \in \mathfrak{g}^*, \quad J(x)(\xi) = \hat{J}(\xi)(x)$$

is called the *momentum map* for ϕ [Marsden and Ratiu (1999); Puta (1993)].

Since ϕ is symplectic, $\phi_{\exp(t\xi)}$ is a one-parameter family of canonical transformations, i.e., $\phi_{\exp(t\xi)}^* \omega = \omega$, hence ξ_M is locally Hamiltonian and not generally Hamiltonian. That is why not every symplectic action has a momentum map. $\phi : G \times M \rightarrow M$ is Hamiltonian iff $\hat{J} : \mathfrak{g} \rightarrow C^k(M, \mathbb{R})$ is a Lie algebra homomorphism.

Let $H : M \rightarrow \mathbb{R}$ be G -invariant, that is $H(\phi_g(x)) = H(x)$ for all $x \in M$ and $g \in G$. Then $\hat{J}(\xi)$ is a constant of motion for dynamics generated by H .

Let ϕ be a symplectic action of G on (M, ω) with the momentum map J . Suppose $H : M \rightarrow \mathbb{R}$ is G -invariant under this action. Then the Noether's Theorem states that J is a constant of motion of H , i.e., $J \circ \phi_t = J$, where ϕ_t is the flow of X_H .

A *Hamiltonian action* is a symplectic action with an Ad^* -equivariant momentum map J , i.e.,

$$J(\phi_g(x)) = Ad_{g^{-1}}^*(J(x)),$$

for all $x \in M$ and $g \in G$.

Let ϕ be a symplectic action of a Lie group G on (M, ω) . Assume that the symplectic form ω on M is exact, i.e., $\omega = d\theta$, and that the action ϕ of G on M leaves the one form $\theta \in M$ invariant. Then $J : M \rightarrow \mathfrak{g}^*$ given

by $(J(x))(\xi) = (i_{\xi_M} \theta)(x)$ is an Ad^* -equivariant momentum map of the action.

In particular, in the case of the cotangent bundle $(M = T^*M, \omega = d\theta)$ of a mechanical configuration manifold M , we can *lift* up an action ϕ of a Lie group G on M to get an action of G on T^*M . To perform this lift, let G act on M by transformations $\phi_g : M \rightarrow M$ and define the *lifted action* to the cotangent bundle by $(\phi_g)_* : T^*M \rightarrow T^*M$ by pushing forward one forms, $(\phi_g)_*(\alpha) \cdot v = \alpha(T\phi_g^{-1}v)$, where $\alpha \in T_q^*M$ and $v \in T_{\phi_g(q)}M$. The lifted action $(\phi_g)_*$ preserves the canonical one form θ on T^*M and the momentum map for $(\phi_g)_*$ is given by

$$J : T^*M \rightarrow \mathfrak{g}^*, \quad J(\alpha_q)(\xi) = \alpha_q(\xi_M(q)).$$

For example, let $M = \mathbb{R}^n$, $G = \mathbb{R}^n$ and let G act on \mathbb{R}^n by translations:

$$\phi : (t, q) \in \mathbb{R}^n \times \mathbb{R}^n \mapsto t + q \in \mathbb{R}^n.$$

Then $\mathfrak{g} = \mathbb{R}^n$ and for each $\xi \in \mathfrak{g}$ we have $\xi_{\mathbb{R}^n}(q) = \xi$.

In case of the group of rotations in \mathbb{R}^3 , $M = \mathbb{R}^3$, $G = SO(3)$ and let G act on \mathbb{R}^3 by $\phi(A, q) = A \cdot q$. Then $\mathfrak{g} \simeq \mathbb{R}^3$ and for each $\xi \in \mathfrak{g}$ we have $\xi_{\mathbb{R}^3}(q) = \xi \times q$.

Let G act transitively on (M, ω) by a Hamiltonian action. Then $J(M) = \{Ad_{g^{-1}}^*(J(x)) \mid g \in G\}$ is a coadjoint orbit.

Now, let (M, ω) be a symplectic manifold, G a Lie group and $\phi : G \times M \rightarrow M$ a Hamiltonian action of G on M with Ad^* -equivariant momentum map $J : M \rightarrow \mathfrak{g}^*$. Let $\mu \in \mathfrak{g}^*$ be a regular value of J ; then $J^{-1}(\mu)$ is a submanifold of M such that $\dim(J^{-1}(\mu)) = \dim(M) - \dim(G)$. Let $G_\mu = \{g \in G \mid Ad_g^* \mu = \mu\}$ be the isotropy subgroup of μ for the coadjoint action. By Ad^* -equivariance, if $x \in J^{-1}(\mu)$ then $\phi_g(x) = J^{-1}(\mu)$ for all $g \in G$, i.e., $J^{-1}(\mu)$ is invariant under the induced G_μ -action and we can form the quotient space $M_\mu = J^{-1}(\mu)/G_\mu$, called the *reduced phase-space* at $\mu \in \mathfrak{g}^*$.

Let (M, ω) be a symplectic $2n$ D manifold and let f_1, \dots, f_k be k functions in involution, i.e., $\{f_i, f_j\}_\omega = 0$, $i, j = 1, \dots, k$. Because the flow of X_{f_i} and X_{f_j} commute, we can use them to define a symplectic action of $G = \mathbb{R}^k$ on M . Here $\mu \in \mathbb{R}^k$ is in the range space of $f_1 \times \dots \times f_k$ and $J = f_1 \times \dots \times f_k$ is the momentum map of this action. Assume that $\{df_1, \dots, df_k\}$ are independent at each point, so μ is a regular value for J . Since G is Abelian, $G_\mu = G$ so we get a symplectic manifold $J^{-1}(\mu)/G$ of dimension $2n - 2k$. If $k = n$ we have *integrable systems*.

For example, let $G = SO(3)$ and $(M, \omega) = (\mathbb{R}^6, \sum_{i=1}^3 dp_i \wedge dq^i)$, and the action of G on \mathbb{R}^6 is given by $\phi : (R, (q, p)) \mapsto (R_q, R_p)$. Then the momentum map is the well known angular momentum and for each $\mu \in \mathfrak{g}^* \simeq \mathbb{R}^3 \mu \neq 0$, $G_\mu \simeq S^1$ and the reduced phase-space (M_μ, ω_μ) is $(T^*\mathbb{R}, \omega = dp_i \wedge dq^i)$, so that $\dim(M_\mu) = \dim(M) - \dim(G) - \dim(G_\mu)$. This reduction is in celestial mechanics called by Jacobi 'the elimination of the nodes'.

The equations of motion: $\dot{f} = \{f, H\}_\omega$ on M reduce to the equations of motion: $\dot{f}_\mu = \{f_\mu, H_\mu\}_{\omega_\mu}$ on M_μ (see [Marsden and Ratiu (1999)]).

3.12.4 Multisymplectic Geometry

Multisymplectic geometry constitutes the general framework for a geometrical, covariant formulation of classical field theory. Here, covariant formulation means that space-like and time-like directions on a given space-time be treated on equal footing. With this principle, one can construct a covariant form of the Legendre transformation which associates to every field variable as many conjugated momenta, the multimomenta, as there are space-time dimensions. These, together with the field variables, those of n D space-time, and an extra variable, the energy variable, span the multiphase-space [Kijowski and Szczyrba (1976)]. For a recent exposition on the differential geometry of this construction, see [Gotay (1991a)]. Multiphase-space, together with a closed, nondegenerate differential $(n + 1)$ -form, the multisymplectic form, is an example of a multisymplectic manifold. Among the achievements of the multisymplectic approach is a geometrical formulation of the relation of infinitesimal symmetries and covariantly conserved quantities (Noether's Theorem), see [León *et al.* (2004)] for a recent review, and [Gotay and Marsden (1992); Forger and Römer (2004)] for a clarification of the improvement techniques ('Belinfante-Rosenfeld formula') of the energy-momentum tensor in classical field theory. Multisymplectic geometry also gives convenient sets of variational integrators for the numerical study of partial differential equations [Marsden *et al.* (1998)].

Since in multisymplectic geometry, the symplectic two-form of classical mechanics is replaced by a closed differential form of higher tensor degree, multivector-fields and differential forms have their natural appearance. (See [Paufler and Römer (2002)] for an interpretation of multivector-fields as describing solutions to field equations infinitesimally.) Multivector-

fields form a graded Lie algebra with the Schouten bracket (see [Kosmann (2004)] for a review and unified viewpoint). Using the multisymplectic $(n + 1)$ -form, one can construct a new bracket for the differential forms, the Poisson forms [Forger *et. al.* (2004)], generalizing a well-known formula for the *Poisson brackets* related to a symplectic two-form. A remarkable fact is that in order to establish a Jacobi identity, the multisymplectic form has to have a potential, a condition that is not seen in symplectic geometry. Further, the admissible differential forms, the Poisson forms, are subject to the rather strong restrictions on their dependence on the multimomentum variables [Gotay (1991b)]. In particular, $(n - 1)$ -forms in this context can be shown to arise exactly from the covariantly conserved currents of Noether symmetries, which allows their pairing with space-like hypersurfaces to yield conserved charges in a geometrical way.

The Hamiltonian, infinite dimensional formulation of classical field theory requires the choice of a space-like hypersurface ('Cauchy surface'), which manifestly breaks the general covariance of the theory at hand. For $(n - 1)$ -forms, the above new bracket reduces to the Peierls-deWitt bracket after integration over the space-like hypersurface [Gotay and Nester (1980)]. With the choice of a hypersurface, a constraint analysis à la Dirac [Henneaux and Teitelboim (1992); Gotay *et. al.* (2004)] can be performed [Landsman (1995)]. Again, the necessary breaking of general covariance raises the need for an alternative formulation of all this [Marsden and Weinstein (1974)]; first attempts have been made to carry out a Marsden-Weinstein reduction [Munteanu *et. al.* (2004)] for multisymplectic manifolds with symmetries. However, not very much is known about how to quantize a multisymplectic geometry, see [Bashkirov and Sardanashvily (2004)] for an approach using a path integral.

So far, everything was valid for field theories of first-order, i.e., where the Lagrangian depends on the fields and their first derivatives. Higher order theories can be reduced to first-order ones for the price of introducing auxiliary fields. A direct treatment would involve higher order jet bundles [Saunders (1989)]. A definition of the covariant Legendre transform and the multiphase-space has been given for this case [Gotay (1991a)].

3.13 Application: Biodynamics-Robotics

Recall from [Ivancevic and Ivancevic (2006)] that modern unified geometrical basis for both human biodynamics and humanoid robotics repre-

sents the *constrained $SE(3)$ -group*, i.e., the so-called *special Euclidean group of rigid-body motions in 3D space* (see, e.g., [Murray *et al.* (1994); Park and Chung (2005)]). In other words, during human movement, in each movable human joint there is an action of a constrained $SE(3)$ -group. Therefore, constrained $SE(3)$ -group represents *general kinematics* of human-like joints. The corresponding nonlinear dynamics problem (resolved mainly for aircraft and spacecraft dynamics) is called the *dynamics on $SE(3)$ -group*, while the associated nonlinear control problem (resolved mainly for general helicopter control) is called the *control on $SE(3)$ -group*.

Recall that the Euclidean $SE(3)$ -group is defined as a semidirect (non-commutative) product of 3D rotations and 3D translations, $SE(3) := SO(3) \triangleright \mathbb{R}^3$ [Murray *et al.* (1994); Park and Chung (2005); Ivancevic and Ivancevic (2006)]). Its most important subgroups are the following:

<i>Subgroup</i>	<i>Definition</i>
$SO(3)$, group of rotations in 3D (a spherical joint)	Set of all proper orthogonal 3×3 – rotational matrices
$SE(2)$, special Euclidean group in 2D (all planar motions)	Set of all 3×3 – matrices: $\begin{bmatrix} \cos \theta & \sin \theta & r_x \\ -\sin \theta & \cos \theta & r_y \\ 0 & 0 & 1 \end{bmatrix}$
$SO(2)$, group of rotations in 2D subgroup of $SE(2)$ – group (a revolute joint)	Set of all proper orthogonal 2×2 – rotational matrices included in $SE(2)$ – group
\mathbb{R}^3 , group of translations in 3D (all spatial displacements)	Euclidean 3D vector space

In the next subsection we give detailed analysis of these subgroups, as well as the total $SE(3)$ -group.

3.13.1 *Muscle-Driven Hamiltonian Biodynamics*

We will develop our Hamiltonian geometry on the configuration biodynamical manifold M in three steps, following the standard symplectic geometry prescription (see subsection 3.12 above):

Step A Find a symplectic *momentum phase-space* (P, ω) .

Recall that a symplectic structure on a smooth manifold M is a nondegenerate closed 2-form ω on M , i.e., for each $x \in M$, $\omega(x)$ is nondegenerate, and $d\omega = 0$.

Let T_x^*M be a cotangent space to M at m . The cotangent bundle T^*M represents a union $\cup_{m \in M} T_x^*M$, together with the standard topology on T^*M and a natural smooth manifold structure, the dimension of which is twice the dimension of M . A 1-form θ on M represents a section $\theta : M \rightarrow T^*M$ of the cotangent bundle T^*M .

$P = T^*M$ is our momentum phase-space. On P there is a nondegenerate symplectic 2-form ω is defined in local joint coordinates $q^i, p_i \in U$, U open in P , as $\omega = dq^i \wedge dp_i$ (' \wedge ' denotes the wedge or exterior product). In that case the coordinates $q^i, p_i \in U$ are called canonical. In a usual procedure the canonical 1-form θ is first defined as $\theta = p_i dq^i$, and then the canonical 2-form ω is defined as $\omega = -d\theta$.

A *symplectic phase-space manifold* is a pair (P, ω) .

Step B Find a *Hamiltonian vector-field* X_H on (P, ω) .

Let (P, ω) be a symplectic manifold. A vector-field $X : P \rightarrow TP$ is called *Hamiltonian* if there is a smooth function $F : P \rightarrow \mathbb{R}$ such that $i_X \omega = dF$ ($i_X \omega$ denotes the *interior product* or *contraction* of the vector-field X and the 2-form ω). X is *locally Hamiltonian* if $i_X \omega$ is closed.

Let the smooth real-valued *Hamiltonian function* $H : P \rightarrow \mathbb{R}$, representing the total biodynamical energy $H(q, p) = T(p) + V(q)$ (T and V denote kinetic and potential energy of the system, respectively), be given in local canonical coordinates $q^i, p_i \in U$, U open in P . The *Hamiltonian vector-field* X_H , condition by $i_{X_H} \omega = dH$, is actually defined via symplectic matrix J , in a local chart U , as

$$X_H = J \nabla H = (\partial_{p_i} H, -\partial_{q^i} H), \quad J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix},$$

where I denotes the $n \times n$ identity matrix and ∇ is the gradient operator.

Step C Find a *Hamiltonian phase-flow* ϕ_t of X_H .

Let (P, ω) be a symplectic phase-space manifold and $X_H = J \nabla H$ a Hamiltonian vector-field corresponding to a smooth real-valued Hamiltonian function $H : P \rightarrow \mathbb{R}$, on it. If a unique one-parameter group of diffeomorphisms $\phi_t : P \rightarrow P$ exists so that $\frac{d}{dt}|_{t=0} \phi_t x = J \nabla H(x)$, it is called the *Hamiltonian phase-flow*.

A smooth curve $t \mapsto (q^i(t), p_i(t))$ on (P, ω) represents an *integral curve* of the Hamiltonian vector-field $X_H = J \nabla H$, if in the local canonical coordinates $q^i, p_i \in U$, U open in P , *Hamiltonian canonical equations* (3.164) hold.

An integral curve is said to be *maximal* if it is not a restriction of an integral curve defined on a larger interval of \mathbb{R} . It follows from the standard

Theorem on the *existence* and *uniqueness* of the solution of a system of ODEs with smooth r.h.s, that if the manifold (P, ω) is Hausdorff, then for any point $x = (q^i, p_i) \in U$, U open in P , there exists a maximal integral curve of $X_H = J\nabla H$, passing for $t = 0$, through point x . In case X_H is complete, i.e., X_H is C^p and (P, ω) is compact, the maximal integral curve of X_H is the Hamiltonian phase-flow $\phi_t : U \rightarrow U$.

The phase-flow ϕ_t is *symplectic* if ω is constant along ϕ_t , i.e., $\phi_t^* \omega = \omega$ ($\phi_t^* \omega$ denotes the *pull-back* of ω by ϕ_t),

iff $\mathcal{L}_{X_H} \omega = 0$

($\mathcal{L}_{X_H} \omega$ denotes the *Lie derivative* of ω upon X_H).

Symplectic phase-flow ϕ_t consists of canonical transformations on (P, ω) , i.e., diffeomorphisms in canonical coordinates $q^i, p_i \in U$, U open on all (P, ω) which leave ω invariant. In this case the *Liouville Theorem* is valid: ϕ_t *preserves* the *phase volume* on (P, ω) . Also, the system's total energy H is conserved along ϕ_t , i.e., $H \circ \phi_t = H$.

Recall that the Riemannian metrics $g = \langle, \rangle$ on the configuration manifold M is a positive-definite quadratic form $g : TM \rightarrow \mathbb{R}$, in local coordinates $q^i \in U$, U open in M , given by (3.139–3.140) above. Given the metrics g_{ij} , the system's Hamiltonian function represents a momentum p —dependent quadratic form $H : T^*M \rightarrow \mathbb{R}$ — the system's kinetic energy $H(p) = T(p) = \frac{1}{2} \langle p, p \rangle$, in local canonical coordinates $q^i, p_i \in U_p$, U_p open in T^*M , given by

$$H(p) = \frac{1}{2} g^{ij}(q, m) p_i p_j, \quad (3.187)$$

where $g^{ij}(q, m) = g_{ij}^{-1}(q, m)$ denotes the *inverse* (contravariant) *material metric tensor*

$$g^{ij}(q, m) = \sum_{\chi=1}^n m_\chi \delta_{rs} \frac{\partial q^i}{\partial x^r} \frac{\partial q^j}{\partial x^s}.$$

T^*M is an *orientable* manifold, admitting the standard *volume form*

$$\Omega_{\omega_H} = \frac{(-1)^{\frac{N(N+1)}{2}}}{N!} \omega_H^N.$$

For Hamiltonian vector-field, X_H on M , there is a base integral curve $\gamma_0(t) = (q^i(t), p_i(t))$ iff $\gamma_0(t)$ is a *geodesic*, given by the one-form *force equation*

$$\dot{p}_i \equiv \dot{p}_i + \Gamma_{jk}^i g^{jl} g^{km} p_l p_m = 0, \quad \text{with} \quad \dot{q}^k = g^{ki} p_i, \quad (3.188)$$

where Γ_{jk}^i denote *Christoffel symbols* of an affine *Levi-Civita connection* on M , defined upon the *Riemannian metric* $g = \langle, \rangle$ by (3.143).

The l.h.s $\dot{\tilde{p}}_i$ of the covariant momentum equation (3.188) represents the *intrinsic* or *Bianchi covariant derivative* of the momentum with respect to time t . Basic relation $\dot{\tilde{p}}_i = 0$ defines the *parallel transport* on T^N , the simplest form of human-motion dynamics. In that case Hamiltonian vector-field X_H is called the *geodesic spray* and its phase-flow is called the *geodesic flow*.

For Earthly dynamics in the gravitational *potential* field $V : M \rightarrow \mathbb{R}$, the Hamiltonian $H : T^*M \rightarrow \mathbb{R}$ (3.187) extends into potential form

$$H(p, q) = \frac{1}{2} g^{ij} p_i p_j + V(q),$$

with Hamiltonian vector-field $X_H = J\nabla H$ still defined by canonical equations (3.164).

A general form of a *driven*, non-conservative Hamiltonian equations reads:

$$\dot{q}^i = \partial_{p_i} H, \quad \dot{p}_i = F_i - \partial_{q^i} H, \quad (3.189)$$

where $F_i = F_i(t, q, p)$ represent any kind of joint-driving *covariant torques*, including active neuro-muscular-like controls, as functions of time, angles and momenta, as well as passive dissipative and elastic joint torques. In the covariant momentum formulation (3.188), the non-conservative Hamiltonian equations (3.189) become

$$\dot{\tilde{p}}_i \equiv \dot{p}_i + \Gamma_{jk}^i g^{jl} g^{km} p_l p_m = F_i, \quad \text{with} \quad \dot{q}^k = g^{ki} p_i.$$

3.13.2 Hamiltonian-Poisson Biodynamical Systems

Recall from subsection 3.12.3.3 above that Hamiltonian-Poisson mechanics is a generalized form of classical Hamiltonian mechanics. Let $(P, \{\})$ be a Poisson manifold and $H \in C^\infty(P, \mathbb{R})$ a smooth real valued function on P . The vector-field X_H defined by

$$X_H(F) = \{F, H\},$$

is the Hamiltonian vector-field with energy function H . The triple $(P, \{\}, H)$ we call the *Hamiltonian-Poisson biodynamical system* (HPBS) [Marsden and Ratiu (1999); Puta (1993); Ivancevic and Pearce (2001a)].

The map $F \mapsto \{F, H\}$ is a derivation on the space $C^\infty(P, \mathbb{R})$, hence it defines a vector-field on P . The map $F \in C^\infty(P, \mathbb{R}) \mapsto X_F \in \mathcal{X}(P)$ is a Lie algebra anti-homomorphism, i.e., $[X_F, X_g] = -X_{\{F, g\}}$.

Let $(P, \{\cdot, \cdot\}, H)$ be a HPBS and ϕ_t the flow of X_H . Then for all $F \in C^\infty(P, \mathbb{R})$ we have the *conservation of energy*:

$$H \circ \phi_t = H,$$

and the *equations of motion in Poisson bracket form*,

$$\frac{d}{dt} (F \circ \phi_t) = \{F, H\} \circ \phi_t = \{F \circ \phi_t, H\},$$

that is, the above Poisson evolution equation (3.179) holds. Now, the function F is constant along the integral curves of the Hamiltonian vector-field X_H iff

$$\{F, H\} = 0.$$

ϕ_t preserves the Poisson structure.

Next we present two main examples of HPBS.

‘Ball-and-Socket’ Joint Dynamics in Euler Vector Form

The dynamics of human body-segments, classically modelled *via* Lagrangian formalism (see [Hatze (1977b); Ivancevic (1991); Ivancevic *et al.* (1995); Ivancevic and Ivancevic (2006)]), may be also prescribed by Euler’s equations of rigid body dynamics. The equations of motion for a free rigid body, described by an observer fixed on the moving body, are usually given by *Euler’s vector equation*

$$\dot{p} = p \times w. \quad (3.190)$$

Here $p, w \in \mathbb{R}^3$, $p_i = I_i w_i$ and I_i ($i = 1, 2, 3$) are the principal moments of inertia, the coordinate system in the segment is chosen so that the axes are principal axes, w is the angular velocity of the body and p is the corresponding angular momentum.

The kinetic energy of the segment is the Hamiltonian function $H : \mathbb{R}^3 \rightarrow \mathbb{R}$ given by [Ivancevic and Pearce (2001a)]

$$H(p) = \frac{1}{2} p \cdot w$$

and is a conserved quantity for (3.190).

The vector space \mathbb{R}^3 is a Lie algebra with respect to the bracket operation given by the usual cross product. The space \mathbb{R}^3 is paired with itself *via* the usual dot product. So if $F : \mathbb{R}^3 \rightarrow \mathbb{R}$, then $\delta F / \delta p = \nabla F(p)$ and the $(-)$ Lie–Poisson bracket $\{F, G\}_-(p)$ is given *via* (3.178) by the triple product

$$\{F, G\}_-(p) = -p \cdot (\nabla F(p) \times \nabla G(p)).$$

Euler’s vector equation (3.190) represents a generalized Hamiltonian system in \mathbb{R}^3 relative to the Hamiltonian function $H(p)$ and the $(-)$ Lie–Poisson bracket $\{F, G\}_-(p)$. Thus the Poisson manifold $(\mathbb{R}^3, \{F, G\}_-(p))$ is defined and the abstract Poisson equation is equivalent to Euler’s equation (3.190) for a body segment and associated joint.

Solitary Model of Muscular Contraction

Recall that the so-called *sliding filament theory of muscular contraction* was developed in 1950s by Nobel Laureate A. Huxley [Huxley and Niedergerke (1954); Huxley (1957)]. At a deeper level, the basis of the molecular model of muscular contraction is represented by oscillations of *Amid I peptide groups* with associated dipole electric momentum inside a spiral structure of myosin filament molecules (see [Davydov (1981); Davydov (1991)]).

There is a simultaneous resonant interaction and strain interaction generating a collective interaction directed along the axis of the spiral. The resonance excitation jumping from one peptide group to another can be represented as an exciton, the local molecule strain caused by the static effect of excitation as a phonon and the resultant collective interaction as a *soliton*.

The simplest model of Davydov’s solitary particle–waves is given by the *nonlinear Schrödinger equation* [Ivancevic and Pearce (2001a)]

$$i\partial_t \psi = -\partial_{x^2} \psi + 2\chi |\psi|^2 \psi, \quad (3.191)$$

for $-\infty < x < +\infty$. Here $\psi(x, t)$ is a smooth complex-valued wave function with initial condition $\psi(x, t)|_{t=0} = \psi(x)$ and χ is a nonlinear parameter. In the linear limit ($\chi = 0$) (3.191) becomes the ordinary Schrödinger equation for the wave function of the free 1D particle with mass $m = 1/2$.

We may define the infinite-dimensional phase-space manifold $\mathcal{P} = \{(\psi, \bar{\psi}) \in S(\mathbb{R}, \mathbf{C})\}$, where $S(\mathbb{R}, \mathbf{C})$ is the Schwartz space of rapidly-decreasing complex-valued functions defined on \mathbb{R}). We define also the

algebra $\chi(\mathcal{P})$ of observables on \mathcal{P} consisting of real-analytic functional derivatives $\delta F/\delta\psi$, $\delta F/\delta\bar\psi \in S(\mathbb{R}, \mathbb{C})$.

The Hamiltonian function $H : \mathcal{P} \rightarrow \mathbb{R}$ is given by

$$H(\psi) = \int_{-\infty}^{+\infty} \left(\left| \frac{\partial\psi}{\partial x} \right|^2 + \chi|\psi|^4 \right) dx$$

and is equal to the total energy of the soliton. It is a conserved quantity for (4.3) (see [Seiler (1995)]).

The Poisson bracket on $\chi(\mathcal{P})$ represents a direct generalization of the classical nD Poisson bracket

$$\{F, G\}_+(\psi) = i \int_{-\infty}^{+\infty} \left(\frac{\delta F}{\delta\psi} \frac{\delta G}{\delta\bar\psi} - \frac{\delta F}{\delta\bar\psi} \frac{\delta G}{\delta\psi} \right) dx. \quad (3.192)$$

It manifestly exhibits skew-symmetry and satisfies Jacobi identity. The functionals are given by $\delta F/\delta\psi = -i\{F, \bar\psi\}$ and $\delta F/\delta\bar\psi = i\{F, \psi\}$. Therefore the algebra of observables $\chi(\mathcal{P})$ represents the Lie algebra and the Poisson bracket is the $(+)$ Lie-Poisson bracket $\{F, G\}_+(\psi)$.

The nonlinear Schrödinger equation (3.191) for the solitary particle-wave is a Hamiltonian system on the Lie algebra $\chi(\mathcal{P})$ relative to the $(+)$ Lie-Poisson bracket $\{F, G\}_+(\psi)$ and Hamiltonian function $H(\psi)$. Therefore the Poisson manifold $(\chi(\mathcal{P}), \{F, G\}_+(\psi))$ is defined and the abstract Poisson evolution equation (3.179), which holds for any smooth function $F : \chi(\mathcal{P}) \rightarrow \mathbb{R}$, is equivalent to equation (3.191).

A more subtle model of soliton dynamics is provided by the *Korteweg-de Vries equation* [Ivancevic and Pearce (2001a)]

$$f_t - 6ff_x + f_{xxx} = 0, \quad (f_x = \partial_x f), \quad (3.193)$$

where $x \in \mathbb{R}$ and f is a real-valued smooth function defined on \mathbb{R} (compare with (3.81) above). This equation is related to the ordinary Schrödinger equation by the inverse scattering method [Seiler (1995); Ivancevic and Pearce (2001a)].

We may define the infinite-dimensional phase-space manifold $\mathcal{V} = \{f \in S(\mathbb{R})\}$, where $S(\mathbb{R})$ is the Schwartz space of rapidly-decreasing real-valued functions \mathbb{R} . We define further $\chi(\mathcal{V})$ to be the algebra of observables consisting of functional derivatives $\delta F/\delta f \in S(\mathbb{R})$.

The Hamiltonian $H : \mathcal{V} \rightarrow \mathbb{R}$ is given by

$$H(f) = \int_{-\infty}^{+\infty} \left(f^3 + \frac{1}{2} f_x^2 \right) dx$$

and gives the total energy of the soliton. It is a conserved quantity for (3.193) (see [Seiler (1995)]).

As a real-valued analogue to (3.192), the (+) Lie-Poisson bracket on $\chi(\mathcal{V})$ is given *via* (3.177) by

$$\{F, G\}_+(f) = \int_{-\infty}^{+\infty} \frac{\delta F}{\delta f} \frac{d}{dx} \frac{\delta G}{\delta f} dx.$$

Again it possesses skew-symmetry and satisfies Jacobi identity. The functionals are given by $\delta F / \delta f = \{F, f\}$.

The Kortevég-de Vries equation (KdV1), describing the behavior of the molecular solitary particle-wave, is a Hamiltonian system on the Lie algebra $\chi(\mathcal{V})$ relative to the (+) Lie-Poisson bracket $\{F, G\}_+(f)$ and the Hamiltonian function $H(f)$. Therefore, the Poisson manifold $(\chi(\mathcal{V}), \{F, G\}_+(f))$ is defined and the abstract Poisson evolution equation (3.179), which holds for any smooth function $F : \chi(\mathcal{V}) \rightarrow \mathbb{R}$, is equivalent to (3.193).

3.13.3 Lie-Poisson Neurodynamics Classifier

A *Lie-Poisson neuro-classifier* is a tensor-field-system $\{\mu\} = (q, p, \omega)$ on a Poisson manifold $(\mathfrak{g}^*, \{F, H(\mu)\}_{\pm}(\mu))$. Like a GBAM neuro-classifier, it consists of continual activation (q, p) -dynamics and self-organized learning ω -dynamics. In this case, both dynamics are defined by “neural activation form” of the abstract Lie-Poisson evolution equation

$$\dot{F} = \{S(F), H(\mu)\}, \quad (3.194)$$

where $S(\cdot) = \tanh(\cdot)$ denotes the sigmoid activation function. A Hamiltonian function $H(\mu)$, representing the total network energy, is given in the form

$$H(\mu) = \frac{1}{2} \omega_{ij} \delta^{ij} + \frac{1}{2} \omega^{ij} \delta_{ij}, \quad (i, j = 1, \dots, n),$$

where δ^{ij} and δ_{ij} are Kronecker tensors, while $\omega^{ij} = \omega^{ij}(q^i)$ and $\omega_{ij} = \omega_{ij}(p_i)$ correspond to the contravariant and covariant components of the functional-coupling synaptic tensor $\omega = \omega(q, p)$, defined respectively by

$$\omega^{ij} = \varepsilon q^i q^j, \quad \omega_{ij} = \tau p_i p_j,$$

with random coefficients ε and τ .

Activation (q, p) -dynamics are given by

$$\dot{q}^i = I^i + \{S(q^i), H(\mu)\}, \quad \dot{p}_i = J_i + \{S(p_i), H(\mu)\},$$

where I^i and J_i represent the two input features.

Two types of self-organized learning ω —dynamics are presented and compared:

Lie–Poisson learning dynamics, in which synaptic update law is given by inhibitory–covariant and excitatory–contravariant form of equation (3.194):

$$\dot{\omega}_{ij} = \{S(\omega_{ij}), H(\mu)\}, \quad \dot{\omega}^{ij} = \{S(\omega^{ij}), H(\mu)\},$$

respectively.

Differential Hebbian learning (see [Kosko (1992)] for details), in both inhibitory–covariant and excitatory–contravariant learning form:

$$\dot{\omega}_{ij} = -\omega_{ij} + \Phi_{ij}(q^i, p_i), \quad \dot{\omega}^{ij} = -\omega^{ij} + \Phi^{ij}(q^i, p_i),$$

with innovations defined in both variance–forms as:

$$\Phi_{ij} = S_i(q^i)S_j(p_j) + \dot{S}_i(q^i)\dot{S}_j(p_j), \quad \Phi^{ij} = S^i(q^i)S^j(p_j) + \dot{S}^i(q^i)\dot{S}^j(p_j).$$

3.13.4 Biodynamical Functors

3.13.4.1 The Covariant Force Functor

Recall (see subsection 2.1.4.3 above) that int the realm of biodynamics the central concept is the *covariant force law*, $F_i = mg_{ij}a^j$ [Ivancevic and Ivancevic (2006)]. In categorical language, it represents the *covariant force functor* \mathcal{F}_* defined by commutative diagram:

$$\begin{array}{ccc}
 TT^*M & \xrightarrow{\mathcal{F}_*} & TTM \\
 \uparrow F_i = \dot{p}_i & & \uparrow a^i = \dot{v}^i \\
 T^*M = \{x^i, p_i\} & & TM = \{x^i, v^i\} \\
 \nwarrow p_i & & \nearrow v^i = \dot{x}^i \\
 & M = \{x^i\} &
 \end{array}$$

which states that the force 1-form $F_i = \dot{p}_i$, defined on the mixed tangent–cotangent bundle TT^*M , causes the acceleration vector–field $a^i = \dot{v}^i$, defined on the second tangent bundle TTM of the configuration manifold M .

The corresponding *contravariant acceleration functor* is defined as its inverse map $\mathcal{F}^* : TTM \rightarrow TT^*M$.

In the following subsections we present several Lie functors, as they are used in modern biodynamical research, all being different formulations of the covariant force law, $F_i = mg_{ij}a^j$, and giving different Lie representations of the fundamental covariant force functor $\mathcal{F}_* : TT^*M \rightarrow TTM$.

3.13.4.2 Lie-Lagrangian Biodynamical Functor

Now we develop the *Lie-Lagrangian biodynamical functor* using a modern, nonlinear formulation of the classical robotics structure (see [Ivancevic (2005b); Ivancevic (2005c)]):

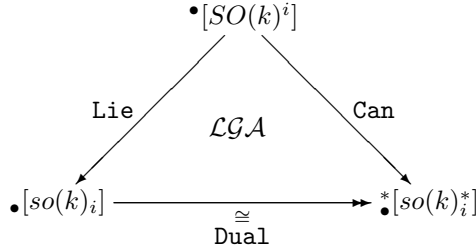
$$\begin{array}{ccccc} \textit{Kinematics} & \rightarrow & \textit{Dynamics} & \rightarrow & \textit{Control} \\ \textit{Lie groups} & \rightarrow & \textit{Exterior Lagrangian} & \rightarrow & \textit{Lie derivative} \end{array}$$

The conservative part of generalized Lagrangian formalism, as used in biodynamics, is derived from Lagrangian conservative energy function. It describes the motion of the conservative skeleton, which is free of control and dissipation. According to the Liouville Theorem, this conservative dynamics is structurally unstable due to the phase-space spreading effect, caused by the growth of entropy (see [Ivancevic (1991); Ivancevic *et al.* (1995); Ivancevic and Snoswell (2001); Ivancevic and Ivancevic (2006)]). The dissipative part is derived from nonlinear dissipative function, and describes quadratic joint dampings, which prevent entropy growth. Its driving part represents equivalent muscular torques F_i acting in all DOF (or just in active joints, as used in the affine input control), in the form of force-time and force-velocity signals.

Joint Kinematics

Recall that human joints represented by internal coordinates x^i ($i = 1, \dots, n$), constitute an n D smooth biodynamical configuration manifold M (see Figure 3.6). Now we are going to perform some categorical transformations on the biodynamical configuration manifold M . If we apply the *functor* **Lie** to the *category* $\bullet[SO(k)^i]$ of rotational Lie groups $SO(k)^i$ and their homomorphisms we get the category $\bullet[so(k)_i]$ of corresponding *tangent* Lie algebras $so(k)_i$ and their homomorphisms. If we further apply the *isomorphic* functor **Dual** to the category $\bullet[so(k)_i]$ we get the dual category $\bullet[so(k)_i^*]$ of *cotangent*, or, *canonical* Lie algebras $so(k)_i^*$ and their

homomorphisms. To go directly from $\bullet[SO(k)^i]$ to $\bullet[so(k)_i^*]$ we use the canonical functor **Can** [Ivancevic and Snoswell (2001); Ivancevic (2002); Ivancevic and Beagley (2005); Ivancevic (2005)]. Therefore we have a commutative triangle



Both the tangent algebras $so(k)_i$ and the cotangent algebras $so(k)_i^*$ contain infinitesimal group generators, angular velocities $\dot{x}^i = \dot{x}^{\phi_i}$ in the first case and canonical angular momenta $p_i = p_{\phi_i}$ in the second. As Lie group generators, angular velocities and angular momenta satisfy the respective commutation relations $[\dot{x}^{\phi_i}, \dot{x}^{\psi_i}] = \epsilon_{\theta}^{\phi\psi} \dot{x}^{\theta_i}$ and $[p_{\phi_i}, p_{\psi_i}] = \epsilon_{\phi\psi}^{\theta} p_{\theta_i}$, where the structure constants $\epsilon_{\theta}^{\phi\psi}$ and $\epsilon_{\phi\psi}^{\theta}$ constitute totally antisymmetric third-order tensors.

In this way, the functor $\text{Dual}_G : \text{Lie} \cong \text{Can}$ establishes a geometrical duality between kinematics of angular velocities \dot{x}^i (involved in Lagrangian formalism on the tangent bundle of M) and that of angular momenta p_i (involved in Hamiltonian formalism on the cotangent bundle of M). This is analyzed below. In other words, we have two functors **Lie** and **Can** from a category of Lie groups (of which $\bullet[SO(k)^i]$ is a subcategory) into a category of their Lie algebras (of which $\bullet[so(k)_i]$ and $\bullet[so(k)_i^*]$ are subcategories), and a natural equivalence (functor isomorphism) between them defined by the functor Dual_G . (As angular momenta p_i are in a bijective correspondence with angular velocities \dot{x}^i , every component of the functor Dual_G is invertible.)

Applying the functor **Lie** to the biodynamical configuration manifold M (Figure 3.6), we get the product-tree of the same anthropomorphic structure, but having tangent Lie algebras $so(k)_i$ as vertices, instead of the groups $SO(k)^i$. Again, applying the functor **Can** to M , we get the product-tree of the same anthropomorphic structure, but this time having cotangent Lie algebras $so(k)_i^*$ as vertices.

The functor **Lie** defines the *second-order Lagrangian formalism on the tangent bundle TM* (i.e., the *velocity phase-space manifold*) while the func-

tor **Can** defines the *first-order canonical Hamiltonian formalism on the cotangent bundle* T^*M (i.e., the *momentum phase-space manifold*). As these two formalisms are related by the isomorphic functor **Dual**, they are equivalent. In this section we shall follow the Lagrangian functor **Lie**, using the powerful formalism of exterior differential systems and integral variational principles [Griffiths (1983); Choquet-Bruhat and DeWitt-Morete (1982)]. For the parallel, Hamiltonian treatment along the functor **Can**, more suitable for chaos theory and stochastic generalizations, see [Ivancevic and Snoswell (2001); Ivancevic (2002)].

Exterior Lagrangian Dynamics

Let $\Omega^p(M) = \sum \omega_I dx^I$ denote the space of differential p -forms on M . That is, if multi-index $I \subset \{1, \dots, n\}$ is a subset of p elements then we have a p -form $dx^I = dx_1^{i_1} \wedge dx_2^{i_2} \wedge \dots \wedge dx_p^{i_p}$ on M . We define the exterior derivative on M as $d\omega = \sum \frac{\partial \omega^I}{\partial x_p} dx_p \wedge dx^I$ (compare with (5.8) above).

Now, from exterior differential systems point of view (see subsection 3.6.2 above as well as [Griffiths (1983)]), human-like motion represents an n DOF neuro-musculo-skeletal system Ξ , evolving in time on its n D configuration manifold M , (with local coordinates x^i , $i = 1, \dots, n$) as well as on its tangent bundle TM (with local coordinates $(x^i; \dot{x}^i)$).

For the system Ξ we will consider a *well-posed variational problem* $(I, \omega; \varphi)$, on an associated $(2n+1)$ -D jet space $X = J^1(\mathbb{R}, M) \cong \mathbb{R} \times TM$, with local canonical variables $(t; x^i; \dot{x}^i)$.

Here, (I, ω) is called a *Pfaffian exterior differential system* on X (see [Griffiths (1983)]), given locally as

$$\begin{cases} \theta^i = dx^i - \dot{x}^i \omega = 0 \\ \omega \equiv dt \neq 0 \end{cases}, \quad (3.195)$$

with the structure equations

$$d\theta^i = -d\dot{x}^i \wedge \omega.$$

Integral manifolds $N \in J^1(\mathbb{R}, M)$ of the Pfaffian system (I, ω) are locally one-jets $t \rightarrow (t, x(t), \dot{x}(t))$ of curves $x = x(t) : \mathbb{R} \rightarrow M$.

φ is a 1-form

$$\varphi = L\omega, \quad (3.196)$$

where $L = L(t, x, \dot{x})$ is the system's *Lagrangian* function defined on X , having both coordinate and velocity partial derivatives, respectively denoted

by

$$L_{x^i} \equiv \partial_{x^i} L, \text{ and } L_{\dot{x}^i} \equiv \partial_{\dot{x}^i} L.$$

A variational problem $(I, \omega; \varphi)$ is said to be *strongly non-degenerate*, or *well-posed* [Griffiths (1983)], if the determinant of the matrix of mixed velocity partials of the Lagrangian is positive definite, i.e.,

$$\det \|L_{\dot{x}^i \dot{x}^j}\| > 0.$$

The *extended Pfaffian system*

$$\begin{cases} \theta^i = 0 \\ dL_{\dot{x}^i} - L_{x^i} \omega = 0 \\ \omega \neq 0 \end{cases}.$$

generates classical *Euler-Lagrangian equations*

$$\frac{d}{dt} L_{\dot{x}^i} = L_{x^i}, \quad (3.197)$$

describing the *control-free, dissipation-free, conservative skeleton dynamics*.

If an *integral manifold* N satisfies the Euler-Lagrangian equations (3.197) of a well-posed variational problem on X then

$$\frac{d}{dt} \left(\int_{N_t} \varphi \right)_{t=0} = 0$$

for any *admissible variation* $N_t \in N$ that satisfies the *endpoint conditions* $\omega = \theta^i = 0$.

Theorem: *Under the above conditions, both the Lagrangian dynamics with initial conditions*

$$\begin{cases} \frac{d}{dt} L_{\dot{x}^i} = L_{x^i} \\ x(t_0) = x_0, \quad \dot{x}(t_0) = \dot{x}_0 \end{cases}$$

and the Lagrangian dynamics with endpoint conditions

$$\begin{cases} \frac{d}{dt} L_{\dot{x}^i} = L_{x^i} \\ x(t_0) = x_0, \quad x(t_1) = x_1 \end{cases}$$

have unique solutions. For the proof, see [Griffiths (1983)].

Now, if M is a smooth Riemannian manifold, its metric $g = \langle \cdot, \cdot \rangle$ is locally given by a positive definite quadratic form

$$ds^2 = g_{ij}(x) dx^i dx^j, \quad (3.198)$$

where the metric tensor g_{ij} is a C^∞ symmetric matrix $g(x) = \|g_{ij}(x)\|$.

Kinetic energy of the system Ξ is a function $T = T(x, \dot{x})$ on the tangent bundle TM , which induces a positive definite quadratic form in each fibre $T_x M \subset TM$. In local coordinates, it is related to the Riemannian metric (3.198) by: $T\omega^2 = \frac{1}{2} ds^2$.

If potential energy of the system Ξ is a function $U = U(x)$ on M , then the autonomous Lagrangian is defined as $L(x, \dot{x}) = T(x, \dot{x}) - U(x)$, i.e., kinetic minus potential energy.

The condition of well-posedness is satisfied, as

$$\det \|L_{\dot{x}^i \dot{x}^j}\| = \det \|g_{ij}(x)\| > 0.$$

Now, the *covariant Euler-Lagrangian equations* (3.197) expand as

$$\frac{d}{dt} (g_{ij}(x(t)) \dot{x}^j(t)) = \frac{1}{2} (\partial_{x^i} g_{jk}(x(t)) \dot{x}^j(t) \dot{x}^k(t) - F_i(x(t))), \quad (3.199)$$

where $F_i(x(t)) = \frac{\partial U(x(t))}{\partial \dot{x}^i}$ denote the *gradient force 1-forms*.

Letting $\|g^{ij}(x)\|$ be the inverse matrix to $\|g_{ij}(x)\|$ and introducing the *Christoffel symbols*

$$\Gamma_{jk}^i = g^{il} \Gamma_{jkl}, \quad \Gamma_{jkl} = \frac{1}{2} (\partial_{x^j} g_{kl} + \partial_{x^k} g_{jl} - \partial_{x^l} g_{jk})$$

the equations (3.199) lead to the classical *contravariant form* (see [Ivancevic (1991); Ivancevic and Pearce (2001b); Ivancevic and Ivancevic (2006)])

$$\ddot{x}^i(t) + \Gamma_{jk}^i(x(t)) \dot{x}^j(t) \dot{x}^k(t) = -F^i(x(t)), \quad (3.200)$$

where $F^i(x(t)) = g^{ij}(x) \frac{\partial U(x(t))}{\partial \dot{x}^j}$ denote the *gradient force vector-fields*.

The above Theorem implies that both the Lagrangian dynamics with initial conditions

$$\begin{cases} \ddot{x}^i(t) + \Gamma_{jk}^i(x(t)) \dot{x}^j(t) \dot{x}^k(t) = -F^i(x(t)) \\ x(t_0) = x_0, \quad \dot{x}(t_0) = \dot{x}_0 \end{cases} \quad (3.201)$$

and the Lagrangian dynamics with endpoint conditions

$$\begin{cases} \ddot{x}^i(t) + \Gamma_{jk}^i(x(t)) \dot{x}^j(t) \dot{x}^k(t) = -F^i(x(t)) \\ x(t_0) = x_0, \quad x(t_1) = x_1 \end{cases} \quad (3.202)$$

have unique solutions. We consider the system (3.201) to be the valid basis of human-like dynamics, and the system (3.202) to be the valid basis of the finite biodynamics control.

Now, recall that any smooth n -manifold M induces an n -category $\Pi_n(M)$, its fundamental n -groupoid. In $\Pi_n(M)$, 0-cells are *points* in M ; 1-cells are *paths* in M (i.e., parameterized smooth maps $f : [0, 1] \rightarrow M$); 2-cells are *smooth homotopies* (denoted by \simeq) of *paths* relative to endpoints (i.e., parameterized smooth maps $h : [0, 1] \times [0, 1] \rightarrow M$); 3-cells are *smooth homotopies of homotopies* of paths in M (i.e., parameterized smooth maps $j : [0, 1] \times [0, 1] \times [0, 1] \rightarrow M$). Categorical *composition* is defined by *pasting* paths and homotopies, which gives the *recursive homotopy dynamics* (see below).

On the other hand, to describe the biodynamical realism, we have to generalize (3.200), so to include any other type of *external* contravariant forces (including excitation and contraction dynamics of muscular-like actuators, as well as nonlinear dissipative joint forces) to the r.h.s of (3.200). In this way, we get the *general form of contravariant Lagrangian dynamics*

$$\ddot{x}^i(t) + \Gamma_{jk}^i(x(t)) \dot{x}^j(t) \dot{x}^k(t) = \mathcal{F}^i(t, x(t), \dot{x}(t)), \quad (3.203)$$

or, in exterior, *covariant* form

$$\frac{d}{dt} L_{\dot{x}^i} - L_{x^i} = \mathcal{F}_i(t, x(t), \dot{x}(t)). \quad (3.204)$$

Recursive homotopy dynamics:

0-cell : $x_0 \bullet \quad x_0 \in M; \quad$ in the higher cells below: $t, s \in [0, 1];$

1-cell : $x_0 \bullet \xrightarrow{f} \bullet x_1 \quad f : x_0 \simeq x_1 \in M,$

$f : [0, 1] \rightarrow M, f : x_0 \mapsto x_1, x_1 = f(x_0), f(0) = x_0, f(1) = x_1;$

e.g., linear path: $f(t) = (1-t)x_0 + tx_1; \quad$ or

e.g., Euler-Lagrangian f -dynamics with endpoint conditions $(x_0, x_1) :$

$\frac{d}{dt} f_{\dot{x}^i} = f_{x^i}, \quad \text{with} \quad x(0) = x_0, \quad x(1) = x_1, \quad (i = 1, \dots, n);$

$$2 - \text{cell} : x_0 \bullet \begin{array}{c} \xrightarrow{f} \\ \Downarrow h \\ \xrightarrow{g} \end{array} \bullet x_1 \quad h : f \simeq g \in M,$$

$$h : [0, 1] \times [0, 1] \rightarrow M, \quad h : f \mapsto g, \quad g = h(f(x_0)),$$

$$h(x_0, 0) = f(x_0), \quad h(x_0, 1) = g(x_0), \quad h(0, t) = x_0, \quad h(1, t) = x_1$$

$$\text{e.g., linear homotopy: } h(x_0, t) = (1 - t)f(x_0) + tg(x_0); \quad \text{or}$$

e.g., homotopy between two Euler–Lagrangian (f, g) – dynamics with the same endpoint conditions (x_0, x_1) :

$$\frac{d}{dt} f_{\dot{x}^i} = f_{x^i}, \quad \text{and} \quad \frac{d}{dt} g_{\dot{x}^i} = g_{x^i} \quad \text{with} \quad x(0) = x_0, \quad x(1) = x_1;$$

$$3 - \text{cell} : x_0 \bullet \begin{array}{c} \xrightarrow{f} \\ \left(\begin{array}{c} \xrightarrow{h} \\ \Downarrow j \\ \xrightarrow{g} \end{array} \right) i \\ \xrightarrow{g} \end{array} \bullet x_1 \quad j : h \simeq i \in M,$$

$$j : [0, 1] \times [0, 1] \times [0, 1] \rightarrow M, \quad j : h \mapsto i, \quad i = j(h(f(x_0)))$$

$$j(x_0, t, 0) = h(f(x_0)), \quad j(x_0, t, 1) = i(f(x_0)),$$

$$j(x_0, 0, s) = f(x_0), \quad j(x_0, 1, s) = g(x_0),$$

$$j(0, t, s) = x_0, \quad j(1, t, s) = x_1$$

$$\text{e.g., linear composite homotopy: } j(x_0, t, s) = (1 - t)h(f(x_0)) + ti(f(x_0));$$

or, homotopy between two homotopies between above two Euler–

Lagrangian (f, g) – dynamics with the same endpoint conditions (x_0, x_1) .

3.13.4.3 Lie–Hamiltonian Biodynamical Functor

The three fundamental and interrelated obstacles facing any researcher in the field of human–like musculo–skeletal dynamics, could be identified as [Ivancevic and Snoswell (2001)]:

- (1) *Deterministic chaos*,
- (2) *Stochastic forces*, and
- (3) *Imprecision of measurement* (or estimation) of the system numbers (SN): inputs, parameters and initial conditions.

Recall that the deterministic chaos is manifested as an irregular and unpredictable time evolution of purely deterministic nonlinear systems. If a nonlinear system is started twice, from slightly different initial conditions, its time evolution differs exponentially, while in case of a linear system, the difference in time evolution is linear.

Again, recall that the stochastic dynamics is based on the concept of *Markov process*⁷, which represents the probabilistic analogue to the deterministic dynamics. The property of a *Markov chain* of prime importance for human-motion dynamics is the existence of an *invariant distribution of states*: we start with an initial state x_0 whose absolute probability is 1. Ultimately the states should be distributed according to a specified distribution.

Recall that *Brownian dynamics* represents the phase-space trajectories of a collection of particles that individually obey *Langevin rate equations* (see [Gardiner (1985)]) in the field of force (i.e., the particles interact with each other via some deterministic force). For one free particle the Langevin equation of motion is given by

$$m\dot{v} = R(t) - \beta v,$$

where m denotes the mass of the particle and v its velocity. The r.h.s represents the coupling to a *heat bath*; the effect of the random force $R(t)$ is to heat the particle. To balance overheating (on the average), the particle is subjected to *friction* β .

Noe, between pure deterministic (in which all DOF of the system in consideration are explicitly taken into account, leading to classical dynamical equations like Hamiltonian) and pure stochastic dynamics (Markov process), there is so-called *hybrid dynamics*, particularly the Brownian dynamics, in which some of DOF are represented only through their *stochastic influence* on others.

System theory and artificial intelligence have long investigated the topic of *uncertainty* in measurement, modelling and simulation. Research in artificial intelligence has enriched the spectrum of available techniques to deal with uncertainty by proposing a theory of possibility, based on the theory of fuzzy sets (see [Yager (1987); Dubois and Prade (1980); Cox (1992); Cox (1994)]). The field of qualitative reasoning and simulation [Berleant and Kuipers (1992)] is also interested in modelling in-

⁷Recall that the Markov process is characterized by a *lack of memory*, i.e., the statistical properties of the immediate future are uniquely determined by the present, regardless of the past (see [Gardiner (1985)]).

completely known systems where qualitative values are expressed by intervals. However, qualitative simulation techniques reveal a low predictive power in presence of complex models. In this section we have combined qualitative and quantitative methods, in spirit of [Bontempi (1995); Ivancevic and Snoswell (2001)].

In this section we will deal with the general biodynamics from the point of view that mathematically and logically approaches a *general theory of systems*, i.e., that *makes the unique framework* for both linear and nonlinear, discrete and continuous, deterministic and stochastic, crisp and fuzzy, SISO and MIMO-systems, and *generalizes the robot dynamics* elaborated in the literature (see [Vukobratovic (1970); Vukobratovic *et al.* (1970); Vukobratovic and Stepanenko (1972); Vukobratovic and Stepanenko (1973); Vukobratovic (1975); Igarashi and Nogai (1992); Hurmuzlu (1993); Shih *et al.* (1993); Shih and Klein (1993)]), including all necessary DOF to match the physiologically realistic human-like motion. Yet, we wish to *avoid* all the mentioned fundamental system *obstacles*. To achieve this goal we have formulated the general *biodynamics functor machine*, covering a union of the three intersected frameworks:

- (1) Muscle-driven, dissipative, Hamiltonian (nonlinear, both discrete and continuous) MIMO-system;
- (2) Stochastic forces (including dissipative fluctuations and ‘Master’ jumps); and
- (3) Fuzzy system numbers.

The Abstract Functor Machine

In this subsection we define the *abstract functor machine* [Ivancevic and Snoswell (2001)] (compare with [Anderson *et al.* (1976)]) by a two-step generalization of the Kalman’s modular theory of linear MIMO-systems [Kalman *et al.* (1969); Kalman (1960)]. The first generalization puts the Kalman’s theory into the category **Vect** of vector spaces and linear operators (see [MacLane (1971)] for technical details about categorical language), thus formulating the unique, categorical formalism valid both for the discrete- and continuous-time MIMO-systems.

We start with the unique, *continual-sequential state equation*

$$\dot{x}(t+1) = Ax(t) + Bu(t), \quad y(t) = Cx(t), \quad (3.205)$$

where the nD vector spaces of *state* $X \ni x$, *input* $U \ni u$, and *output* $Y \ni y$ have the corresponding linear operators, respectively $A : X \rightarrow X$,

$B : U \rightarrow X$, and $C : X \rightarrow Y$. The modular system theory comprises the *system dynamics*, given by a pair (X, A) , together with a *reachability map* $e : U \rightarrow X$ of the pair (B, A) , and an *observability map* $m : X \rightarrow Y$ of the pair (A, C) . If the reachability map e is surjection the system dynamics (X, A) is called *reachable*; if the observability map m is injection the system dynamics (X, A) is called *observable*. If the system dynamics (X, A) is both reachable and observable, a *composition* $r = m \circ e : U \rightarrow Y$ defines the *total system's response*, which is given by solution of equation (3.205). If the unique solution to the continual-sequential state equation exists, it gives the answer to the (minimal) *realization problem*: find the system S that realizes the given response $r = m \circ e : U \rightarrow Y$ (in the smallest number of discrete states and in the shortest time).

In categorical language, the system dynamics in the category **Vect** is a pair (X, A) , where $X \in \text{Ob}(\mathbf{Vect})$ is an object in **Vect** and $A : X \rightarrow X \in \text{Mor}(\mathbf{Vect})$ is a **Vect**-morphism. A *decomposable system* in **Vect** is such a sextuple $S \equiv (X, A, U, B, Y, C)$ that (X, A) is the system dynamics in **Vect**, a **Vect**-morphism $B : U \rightarrow X$ is an *input map*, and a **Vect**-morphism $C : X \rightarrow Y$ is an *output map*. Any object in **Vect** is characterized by mutually *dual*⁸ notions of its *degree* (a number of its input morphisms) and its *codegree* (a number of its output morphisms). Similarly, any decomposable system S in **Vect** has a *reachability map* given by an epimorphism $e = A \circ B : U \rightarrow X$ and its dual *observability map* given by a monomorphism $m = C \circ A : X \rightarrow Y$; their composition $r = m \circ e : U \rightarrow Y$ in **Mor(Vect)** defines the total system's response in **Vect** given by the unique solution of the continual-sequential state equation (3.205).

The second generalization gives an extension of the continual-sequential MIMO-system theory: from the linear category **Vect** – to an arbitrary nonlinear category \mathcal{K} . We do this extension (see [Ivancevic and Snoswell (2001)]) by formally applying the *action* of the *nonlinear process-functor* $\mathcal{F} : \mathcal{K} \Rightarrow \mathcal{K}$ on the decomposable system $S \equiv (X, A, U, B, Y, C)$ in **Vect**. Under the action of the process functor \mathcal{F} the linear system dynamics (X, A) in **Vect** transforms into a nonlinear \mathcal{F} -dynamics $(\mathcal{F}[X], \mathcal{F}[A])$ in \mathcal{K} , creating the *functor machine* in \mathcal{K} represented by a *nonlinear decomposable system* $\mathcal{F}[S] \equiv (\mathcal{F}[X], \mathcal{F}[A], \mathcal{F}[U], \mathcal{F}[B], \mathcal{F}[Y], \mathcal{F}[C])$. The reachability map transforms into the *input process* $\mathcal{F}[e] = \mathcal{F}[A] \circ \mathcal{F}[B] : \mathcal{F}[U] \rightarrow \mathcal{F}[X]$, while its dual, observability map transforms into the *output process* $\mathcal{F}[m] =$

⁸Recall that in categorical language *duality* means reversing the (arrows of) morphisms; the knowledge of one of the two mutually dual terms automatically implies the knowledge of the other.

$\mathcal{F}[C] \circ \mathcal{F}[A] : \mathcal{F}[X] \rightarrow \mathcal{F}[Y]$. In this way the total response of the linear system $r = m \circ e : U \rightarrow Y$ in $\mathbf{Mor}(\mathbf{Vect})$ transforms into the *nonlinear system behavior* $\mathcal{F}[r] = \mathcal{F}[m] \circ \mathcal{F}[e] : \mathcal{F}[U] \rightarrow \mathcal{F}[Y]$ in $\mathbf{Mor}(\mathcal{K})$. Obviously, $\mathcal{F}[r]$, if exists, is given by a nonlinear \mathcal{F} –transform of the linear state equation (3.205).

The purpose of this section is to formulate a nonlinear \mathcal{F} –transform for the linear state equation (3.205) for biodynamics, i.e., the biodynamics functor machine. In subsequent sections we give a three–step development of a fuzzy–stochastic–Hamiltonian formulation for the biodynamics functor machine $\mathcal{F}[S]$, with a corresponding nonlinear system behavior $\mathcal{F}[r]$.

Muscle–Driven, Dissipative, Hamiltonian Biodynamics

In this subsection we choose the functor **Can**, as the first–order Hamiltonian formalism is more suitable for both stochastic and fuzzy generalizations to follow. Recall that the general deterministic Hamiltonian biodynamics, representing the canonical functor $\mathbf{Can} : \mathcal{S}^\bullet[SO(n)^i] \Rightarrow \mathcal{S}_\bullet[so(n)_i^*]$, is given by dissipative, driven δ –Hamiltonian equations,

$$\dot{q}^i = \frac{\partial H}{\partial p_i} + \frac{\partial R}{\partial p_i}, \quad (3.206)$$

$$\dot{p}_i = F_i - \frac{\partial H}{\partial q^i} + \frac{\partial R}{\partial q^i}, \quad (3.207)$$

$$q^i(0) = q_0^i, \quad p_i(0) = p_i^0, \quad (3.208)$$

including *contravariant* equation (3.206) – the *velocity vector–field*, and *covariant* equation (3.207) – the *force 1–form*, together with initial joint angles and momenta (3.208). Here $(i = 1, \dots, N)$, and $R = R(q, p)$ denotes the Raileigh nonlinear (biquadratic) dissipation function, and $F_i = F_i(t, q, p)$ are covariant driving torques of *equivalent muscular actuators*, resembling muscular excitation and contraction dynamics in rotational form.

The velocity vector–field (3.206) and the force 1–form (3.207) together define the generalized Hamiltonian vector–field X_H , which geometrically represents the *section* of the momentum phase–space manifold T^*M , which is itself the cotangent bundle of the biodynamical configuration manifold M ; the Hamiltonian (total energy) function $H = H(q, p)$ is its generating function.

As a Lie group, the configuration manifold M is Hausdorff [Abraham *et al.* (1988); Marsden and Ratiu (1999); Postnikov (1986)]. Therefore, for $x = (q^i, p_i) \in U_p$, U_p open in T^*M , there exists a unique one–parameter

group of diffeomorphisms $\phi_{\delta_t} : T^*M \rightarrow T^*M$, the generalized deterministic δ –Hamiltonian phase–flow

$$\begin{aligned} \phi_{\delta_t} : G_1 \times T^*M &\rightarrow T^*M : (p(0), q(0)) \mapsto (p(t), q(t)), \\ (\phi_{\delta_t} \circ \phi_{\delta_s} &= \phi_{\delta_{t+s}}, \quad \phi_{\delta_0} = \text{identity}), \end{aligned} \quad (3.209)$$

given by (3.206–3.208) such that

$$\frac{d}{dt} \Big|_{t=0} \phi_{\delta_t} x = J \nabla H(x).$$

The δ –Hamiltonian system (3.206–3.208), with its δ –Hamiltonian phase–flow ϕ_{δ_t} (3.209), i.e., the canonical functor **Can**, represents our first, continual–deterministic model for the biodynamics functor machine $\mathcal{F}[S]$ with the nonlinear system behavior $\mathcal{F}[r]$. In the two subsequent sections we generalize this model to include discrete stochastic forces and fuzzy SN.

Stochastic–Lie–Hamiltonian Biodynamical Functor

In terms of the Markov stochastic process, we can interpret the deterministic δ –Hamiltonian biodynamical system (3.206–3.208) as deterministic drift corresponding to the *Liouville equation*. Thus, we can naturally (in the sense of Langevin) add the covariant vector $\sigma_i(t)$ of stochastic forces (diffusion fluctuations and discontinuous–Master jumps) $\sigma_i(t) = B_{ij}[q^i(t), t] dW^j(t)$ to the canonical force equation. In this way we get *stochastic* σ –Hamiltonian biodynamical system, a *stochastic transformation* **Stoch**[**Can**] of the canonical functor **Can**,

$$dq^i = \left(\frac{\partial H}{\partial p_i} + \frac{\partial R}{\partial p_i} \right) dt, \quad (3.210)$$

$$dp_i = \left(F_i - \frac{\partial H}{\partial q^i} + \frac{\partial R}{\partial q^i} \right) dt + \sigma_i(t), \quad (3.211)$$

$$\sigma_i(t) = B_{ij}[q^i(t), t] dW^j(t), \quad q^i(0) = q_0^i, \quad p_i(0) = p_i^0.$$

In our low–dimensional example–case of symmetrical 3D load–lifting, the velocity and force σ –Hamiltonian biodynamics equations (3.210–3.211)

become

$$dq^i = \left(p_i \left\{ [J_i]^{-1} + \left[m_i \left(\sum_{j=1}^i L_j \cos q^j \right)^2 \right]^{-1} \right\} + \frac{\partial R}{\partial p_i} \right) dt,$$

$$dp_i = B_{ij}[q^i(t), t] dW^j(t) + \left(F_i - g \sum_{j=i}^{10-i} L_j m_j \sin q^j \right. \\ \left. - \sum_{j=i}^{10-i} L_j \sin q^j p_i p_j \left[m_i \left(\sum_{k=1}^i L_k \cos q^k \right)^3 \right]^{-1} + \frac{\partial R}{\partial q^i} \right) dt.$$

Recall that Ito quadratic cotangent bundle I^*Q^N is defined as a Whitney sum

$$I^*Q^N = T^*Q^N \oplus SQ^N,$$

where SQ^N corresponds to *stochastic tensor bundle*, whose elements are 2nd-order tensor-fields composed of continual diffusion fluctuations and discontinuous jumps at every point of the manifold Q^N . On I^*Q^N is defined a non-degenerate, stochastic 2-form α which is closed, i.e., $d\alpha = 0$, and exact, i.e., $\alpha = d\beta$, where 1-form β represents a section $\beta : Q^N \rightarrow I^*Q^N$ of the Ito bundle I^*Q^N .

Now, the stochastic Hamiltonian vector-field Ξ_H represents a section $\Xi_H : Q^N \rightarrow IQ^N$ of the *Ito quadratic tangent bundle* IQ^N , also defined as a Whitney sum

$$IQ^N = TM \oplus SQ^N.$$

The quadratic character of Ito stochastic fibre-bundles corresponds to the second term (trace of the 2nd-order tensor-field) of associate *stochastic Taylor expansion* (see [Elworthy (1982); Mayer (1981)]).

Through stochastic σ -Hamiltonian biodynamical system (3.210–3.211), the deterministic δ -Hamiltonian phase-flow ϕ_{δ_t} (3.209), extends into stochastic σ -Hamiltonian phase-flow ϕ_{σ_t}

$$\phi_{\sigma_t} : G_1 \times I^*M \rightarrow I^*M : (p(0), q(0)) \mapsto (p(t), q(t)), \quad (3.212)$$

$$(\phi_{\sigma_t} \circ \phi_{\sigma_s} = \phi_{\sigma_{t+s}}, \quad \phi_{\sigma_0} = \text{identity}),$$

where I^*M denotes *Ito quadratic cotangent bundle* (see [Elworthy (1982); Mayer (1981)]) of biodynamical configuration manifold M .

Besides the σ -Hamiltonian phase-flow ϕ_{σ_t} (3.212), including N individual random-phase trajectories, we can also define (see [Elworthy (1982)]) an *average* or *mean* $\langle \sigma \rangle$ -Hamiltonian flow $\langle \phi \rangle_{\sigma_t}$

$$\begin{aligned} \langle \phi \rangle_{\sigma_t} : G_1 \times I^*M &\rightarrow I^*M : (\langle p(0) \rangle, \langle q(0) \rangle) \mapsto (\langle p(t) \rangle, \langle q(t) \rangle), \\ (\langle \phi \rangle_{\sigma_t} \circ \langle \phi \rangle_{\sigma_s} &= \langle \phi \rangle_{\sigma_{t+s}}, \quad \langle \phi \rangle_{\sigma_0} = \text{identity}), \end{aligned}$$

which stochastically corresponds to the trajectory of the center of mass in the human-like dynamics, approximatively lumbo-sacral spinal $SO(3)$ -joint.

The necessary conditions for existence of a unique *non-anticipating solution* of the σ -Hamiltonian biodynamical system in a fixed time interval are *Lipschitz condition* and *growth condition* (see [Elworthy (1982); Mayer (1981)]). For constructing an approximate solution a simple iterative Cauchy-Euler procedure could be used to calculate (q_{k+1}^i, p_i^{k+1}) from the knowledge of (q_k^i, p_i^k) on the mesh of time points t^k , $k = 1, \dots, s$, by adding discrete δ -Hamiltonian drift-terms $A^i(q_k^i)\Delta t^k$ and $A_i(p_i^k)\Delta t^k$, as well as a stochastic term $B_{ij}(q_i^k, t^k)\Delta W_k^j$.

σ -Hamiltonian biodynamical system (3.210–3.211), with its σ -Hamiltonian phase-flow ϕ_{σ_t} (3.212), i.e., the functor **Stoch**[Can], represents our second, continual-discrete stochastic model for the biodynamics functor machine $\mathcal{F}[S]$ with the nonlinear system behavior $\mathcal{F}[r]$. In the next section we generalize this model once more to include fuzzy SN.

Fuzzy-Stochastic-Lie-Hamiltonian Functor

Generally, a *fuzzy differential equation* model (FDE-model, for short) is a symbolic description expressing a state of incomplete knowledge of the continuous world, and is thus an abstraction of an infinite set of ODEs models. Qualitative simulation (see [Berleant and Kuipers (1992)]) predicts the set of possible behaviors consistent with a FDE model and an initial state. Specifically, as a FDE we consider an ordinary deterministic (i.e., *crisp*) differential equation (CDE) in which some of the parameters (i.e., coefficients) or initial conditions are *fuzzy numbers*, i.e., uncertain and represented in a *possibilistic* form. As a *solution* of a FDE we consider a time evolution of a *fuzzy region of uncertainty* in the system's phase-space, which corresponds to its the *possibility distribution*.

Recall that a fuzzy number is formally defined as a convex, normalized *fuzzy set* [Dubois and Prade (1980); Cox (1992); Cox (1994)]. The concept

of fuzzy numbers is an extension of the notion of real numbers: it encodes approximate quantitative knowledge. It is not probabilistic, but rather a possibilistic distribution. The mathematics of fuzzy numbers is founded on the *extension principle*, introduced by Zadeh [Yager (1987)]. This principle gives a general method for extending standard mathematical concepts in order to deal with fuzzy quantities [Dubois and Prade (1980)].

Let $\Phi : Y^1 \times Y^2 \times \dots \times Y^n \rightarrow Z$ be a deterministic map such that $z = \Phi(y^1, y^2, \dots, y^n)$ for all $z \in Z$, $y^i \in Y^i$. The extension principle allows us to induce from n input fuzzy sets \bar{y}^i on Y^i an output fuzzy set \bar{z} on Z through Φ given by

$$\mu_{\bar{z}}(t) = \sup_{t=\Phi(s^1, \dots, s^n)} \min(\mu_{\bar{y}^1}(s^1), \dots, \mu_{\bar{y}^n}(s^n)),$$

$$\text{or} \quad \mu_{\bar{z}}(t) = 0 \quad \text{if} \quad \Phi^{-1}(t) = \emptyset,$$

where $\Phi^{-1}(t)$ denotes the inverse image of t and $\mu_{\bar{y}^i}$ is the membership function of \bar{y}^i , ($i = 1, \dots, n$).

The extension principle gives a method to calculate the fuzzy value of a fuzzy map but, in practice, its application is not feasible because of the infinite number of computations it would require. The simplest way of efficiently applying the extension principle is in the form of iterative repetition of several crisp Hamiltonian simulations (see [Bontempi (1995); Ivancevic and Snoswell (2001); Pearce and Ivancevic (2003); Pearce and Ivancevic (2004)]), within the range of included fuzzy SN.

Fuzzification of the crisp deterministic δ -Hamiltonian biodynamical system (3.206–3.208) gives the fuzzified μ -Hamiltonian biodynamical system, namely δ -Hamiltonian biodynamical system with fuzzy SN, i.e., the *fuzzy transformation* **Fuzzy**[**Can**] of the canonical functor **Can**

$$\dot{q}^i = \frac{\partial H(q, p, \sigma)}{\partial p_i} + \frac{\partial R}{\partial p_i}, \quad (3.213)$$

$$\dot{p}_i = \bar{F}_i(q, p, \sigma) - \frac{\partial H(q, p, \sigma)}{\partial q^i} + \frac{\partial R}{\partial q^i}, \quad (3.214)$$

$$q^i(0) = \bar{q}_0^i, \quad p_i(0) = \bar{p}_i^0, \quad (i = 1, \dots, N). \quad (3.215)$$

Here $\sigma = \sigma_\mu$ (with $\mu \geq 1$) denote fuzzy sets of conservative parameters (segment lengths, masses and moments of inertia), dissipative joint dampings and actuator parameters (amplitudes and frequencies), while the bar ($\bar{\cdot}$) over a variable (\cdot) denotes the corresponding fuzzified variable.

In our example–case of symmetrical 3D load–lifting, the fuzzified μ –Hamil–tonian biodynamical system (3.213–3.215) becomes

$$\begin{aligned} \dot{q}^i &= p_i \left\{ [\bar{J}_i]^{-1} + \left[\bar{m}_i \left(\sum_{j=1}^i \bar{L}_j \cos q^j \right)^2 \right]^{-1} \right\} + \frac{\partial R}{\partial p_i}, \\ \dot{p}_i &= \bar{F}_i(t, q^i, p_i, \{\sigma\}_\mu) - g \sum_{j=i}^{10-i} \bar{L}_j \bar{m}_j \sin q^j \\ &\quad - \sum_{j=i}^{10-i} \bar{L}_j \sin q^j p_i p_j \left[\bar{m}_i \left(\sum_{k=1}^i \bar{L}_k \cos q^k \right)^3 \right]^{-1} + \frac{\partial R}{\partial q^i} \\ q^i(0) &= \bar{q}_0^i, \quad p_i(0) = \bar{p}_i^0, \quad (i = 1, \dots, 9). \end{aligned}$$

In this way, the crisp δ –Hamiltonian phase–flow ϕ_{δ_t} (3.209) extends into fuzzy–deterministic μ –Hamiltonian phase–flow ϕ_{μ_t}

$$\begin{aligned} \phi_{\mu_t} : G_1 \times T^*M &\rightarrow T^*M : (\bar{p}_i^0, \bar{q}_0^i) \mapsto (p(t), q(t)), \\ (\phi_{\mu_t} \circ \phi_{\mu_s} &= \phi_{\mu_{t+s}}, \quad \phi_{\mu_0} = \text{identity}). \end{aligned}$$

Similarly, fuzzification of crisp stochastic σ –Hamiltonian biodynamical system (3.210–3.211) gives fuzzy–stochastic $[\mu\sigma]$ –Hamiltonian biodynamical system, namely stochastic σ –Hamiltonian biodynamical system with fuzzy SN, i.e., the *fuzzy–stochastic transformation* **Fuzzy[Stoch[Can]]** of the canonical functor **Can**

$$dq^i = \left(\frac{\partial H(q, p, \sigma)}{\partial p_i} + \frac{\partial R}{\partial p_i} \right) dt, \quad (3.216)$$

$$dp_i = B_{ij}[q^i(t), t] dW^j(t) + \left(\bar{F}_i(q, p, \sigma) - \frac{\partial H(q, p, \sigma)}{\partial q^i} + \frac{\partial R}{\partial q^i} \right) dt, \quad (3.217)$$

$$q^i(0) = \bar{q}_0^i, \quad p_i(0) = \bar{p}_i^0. \quad (3.218)$$

In our example–case of symmetrical 3D load–lifting, the velocity and force $[\mu\sigma]$ –Hamiltonian biodynamics equations (3.216–3.217) become

$$\begin{aligned} dq^i &= \left(p_i \left\{ [\bar{J}_i]^{-1} + \left[\bar{m}_i \left(\sum_{j=1}^i \bar{L}_j \cos q^j \right)^2 \right]^{-1} \right\} + \frac{\partial R}{\partial p_i} \right) dt, \\ dp_i &= B_{ij}[q^i(t), t] dW^j(t) + \left(\bar{F}_i(t, q^i, p_i, \{\sigma\}_\mu) - g \sum_{j=i}^{10-i} \bar{L}_j \bar{m}_j \sin q^j \right. \\ &\quad \left. - \sum_{j=i}^{10-i} \bar{L}_j \sin q^j p_i p_j \left[\bar{m}_i \left(\sum_{k=1}^i \bar{L}_k \cos q^k \right)^3 \right]^{-1} + \frac{\partial R}{\partial q^i} \right) dt. \end{aligned}$$

In this way, the crisp stochastic σ –Hamiltonian phase–flow ϕ_{σ_t} (3.212) extends into fuzzy–stochastic $[\mu\sigma]$ –Hamiltonian phase–flow $\phi_{[\mu\sigma]_t}$

$$\begin{aligned} \phi_{[\mu\sigma]_t} : G_1 \times I^*M &\rightarrow I^*M : (\bar{p}_i^0, \bar{q}_0^i) \mapsto (p(t), q(t)), \quad (3.219) \\ (\phi_{[\mu\sigma]_t} \circ \phi_{[\mu\sigma]_s}) &= \phi_{[\mu\sigma]_{t+s}}, \quad \phi_{[\mu\sigma]_0} = \text{identity}. \end{aligned}$$

$[\mu\sigma]$ –Hamiltonian biodynamical system (3.216–3.218), with its phase–flow $\phi_{[\mu\sigma]_t}$ (3.219), i.e., the functor **Fuzzy[Stoch[Can]]**, represents our final, continual–discrete and fuzzy–stochastic model for the biodynamics functor machine $\mathcal{F}[S]$ with the nonlinear system behavior $\mathcal{F}[r]$.

3.13.5 Biodynamical Topology

3.13.5.1 (Co)Chain Complexes in Biodynamics

In this section we present the category of (co)chain complexes, as used in modern biodynamics. The central concept in cohomology theory is the *category* $\mathbf{S}^\bullet(\mathbb{C})$ of *generalized cochain complexes* in an Abelian category \mathbb{C} [Dieudonne (1988)]. The *objects* of the category $\mathbf{S}^\bullet(\mathbb{C})$ are infinite sequences

$$A^\bullet : \dots \longrightarrow A^{n-1} \xrightarrow{d^{n-1}} A^n \xrightarrow{d^n} A^{n+1} \longrightarrow \dots$$

where, for each $n \in \mathbb{Z}$, A^n is an object of \mathbb{C} and d^n a morphism of \mathbb{C} , with the conditions

$$d^{n-1} \circ d^n = 0$$

for every $n \in \mathbb{Z}$. When $A^n = 0$ for $n < 0$, one speaks of *cochain complexes*. The d^n are called *coboundary operators*.

The *morphisms* of the category $\mathbf{S}^\bullet(\mathbb{C})$ are sequences $f^\bullet = (f^n) : A^\bullet \rightarrow B^\bullet$ where, for each $n \in \mathbb{Z}$, $f^n : A^n \rightarrow B^n$ is a morphism of \mathbb{C} , and in the diagram

$$\begin{array}{ccccccc} \dots & \longrightarrow & A^{n-1} & \xrightarrow{d^{n-1}} & A^n & \xrightarrow{d^n} & A^{n+1} \longrightarrow \dots \\ & & \downarrow f^{n-1} & & \downarrow f^n & & \downarrow f^{n+1} \\ \dots & \longrightarrow & B^{n-1} & \xrightarrow{d^{n-1}} & B^n & \xrightarrow{d^n} & B^{n+1} \longrightarrow \dots \end{array} \quad (3.220)$$

all squares are commutative; one says the f^n commute with the coboundary operators. One has $\text{Im } d^{n+1} \subset \text{Ker } d^n \subset A^n$ for every $n \in \mathbb{Z}$; the quotient $H^n(A^\bullet) = \text{Ker } d^n / \text{Im } d^{n+1}$ is called the n th cohomology object of A^\bullet . From (3.220) it follows that there is a morphism

$$H^n(f^\bullet) : H^n(A^\bullet) \rightarrow H^n(B^\bullet)$$

deduced canonically from f^\bullet , and

$$(A^\bullet, f^\bullet) \Rightarrow (H^n(A^\bullet), H^n(f^\bullet))$$

is a *covariant functor* from $\mathbf{S}^\bullet(\mathbb{C})$ to \mathbb{C} .

The *cohomology exact sequence*: if three cochain complexes $A^\bullet, B^\bullet, C^\bullet$ are elements of a short exact sequence of morphisms

$$0 \longrightarrow A^\bullet \longrightarrow B^\bullet \longrightarrow C^\bullet \longrightarrow 0$$

then there exists an infinite sequence of canonically defined morphisms $d^n : H^n(C^\bullet) \rightarrow H^{n-1}(A^\bullet)$ such that the sequence

$$\dots \longrightarrow H^n(A^\bullet) \longrightarrow H^n(B^\bullet) \longrightarrow H^n(C^\bullet) \longrightarrow H^{n-1}(A^\bullet) \longrightarrow \dots$$

is *exact*, that is the *image* of each homomorphism in the sequence is exactly the *kernel* of the next one.

The *dual* to the category $\mathbf{S}^\bullet(\mathbb{C})$ is the *category* of $\mathbf{S}_\bullet(\mathbb{C})$ of *generalized chain complexes*. Its objects and morphisms are get by formal inversion of all arrows and lowering all indices.

Biodynamical (Co)Homologies

Let \mathcal{M}^\bullet denote the Abelian category of cochains, (i.e., p –forms) on the biodynamical configuration manifold M (see Figure 3.7). When $\mathcal{C} = \mathcal{M}^\bullet$, we have the category $\mathcal{S}^\bullet(\mathcal{M}^\bullet)$ of generalized cochain complexes A^\bullet in \mathcal{M}^\bullet , and if $A' = 0$ for $n < 0$ we have a subcategory $\mathcal{S}_{\mathcal{DR}}^\bullet(\mathcal{M}^\bullet)$ of the *de Rham differential complexes* in \mathcal{M}^\bullet

$$A_{\mathcal{DR}}^\bullet : 0 \rightarrow \Omega^0(M) \xrightarrow{d} \Omega^1(M) \xrightarrow{d} \Omega^2(M) \cdots \xrightarrow{d} \Omega^n(M) \xrightarrow{d} \cdots .$$

Here $A' = \Omega^n(M)$ is the vector space over \mathbb{R} of all p –forms ω on M (for $p = 0$ the smooth functions on M) and $d_n = d : \Omega^{n-1}(M) \rightarrow \Omega^n(M)$ is the exterior differential. A form $\omega \in \Omega^n(M)$ such that $d\omega = 0$ is a closed form or n –cocycle. A form $\omega \in \Omega^n(M)$ such that $\omega = d\theta$, where $\theta \in \Omega^{n-1}(M)$, is an exact form or n –coboundary. Let $Z^n(M) = \text{Ker } d$ (resp. $B^n(M) = \text{Im } d$) denote a real vector space of cocycles (resp. coboundaries) of degree n . Since $d_{n+1} \circ d_n = d^2 = 0$, we have $B^n(M) \subset Z^n(M)$. The quotient vector space

$$H_{\mathcal{DR}}^n(M) = \text{Ker } d / \text{Im } d = Z^n(M) / B^n(M)$$

is the *de Rham cohomology group*. The elements of $H_{\mathcal{DR}}^n(M)$ represent equivalence sets of cocycles. Two cocycles ω_1, ω_2 belong to the same equivalence set, or are cohomologous (written $\omega_1 \sim \omega_2$) iff they differ by a coboundary $\omega_1 - \omega_2 = d\theta$. The de Rham’s cohomology class of any form $\omega \in \Omega^n(M)$ is $[\omega] \in H_{\mathcal{DR}}^n(M)$. The de Rham differential complex (1) can be considered as a system of second-order DEs $d^2\theta = 0$, $\theta \in \Omega^{n-1}(M)$ having a solution represented by $Z^n(M) = \text{Ker } d$.

Analogously let \mathcal{M}_\bullet denote the Abelian category of chains on the configuration manifold M . When $\mathcal{C} = \mathcal{M}_\bullet$, we have the category $\mathcal{S}_\bullet(\mathcal{M}_\bullet)$ of generalized chain complexes A_\bullet in \mathcal{M}_\bullet , and if $A_n = 0$ for $n < 0$ we have a subcategory $\mathcal{S}_\bullet^{\mathcal{C}}(\mathcal{M}_\bullet)$ of chain complexes in \mathcal{M}_\bullet .

$$A_\bullet : 0 \leftarrow C^0(M) \xleftarrow{\partial} C^1(M) \xleftarrow{\partial} C^2(M) \cdots \xleftarrow{\partial} C^n(M) \xleftarrow{\partial} \cdots .$$

Here $A_n = C^n(M)$ is the vector space over \mathbb{R} of all finite chains C on the manifold M and $\partial_n = \partial : C^{n+1}(M) \rightarrow C^n(M)$. A finite chain C such that $\partial C = 0$ is an n –cycle. A finite chain C such that $C = \partial B$ is an n –boundary. Let $Z_n(M) = \text{Ker } \partial$ (resp. $B_n(M) = \text{Im } \partial$) denote a real vector space of cycles (resp. boundaries) of degree n . Since $\partial_{n+1} \circ \partial_n =$

$\partial^2 = 0$, we have $B_n(M) \subset Z_n(M)$. The quotient vector space

$$H_n^C(M) = \text{Ker } \partial / \text{Im } \partial = Z_n(M) / B_n(M)$$

is the n -homology group. The elements of $H_n^C(M)$ are equivalence sets of cycles. Two cycles C_1, C_2 belong to the same equivalence set, or are homologous (written $C_1 \sim C_2$), iff they differ by a boundary $C_1 - C_2 = \partial B$. The homology class of a finite chain $C \in C^n(M)$ is $[C] \in H_n^C(M)$.

The dimension of the n -cohomology (resp. n -homology) group equals the n th Betti number b^n (resp. b_n) of the manifold M . *Poincaré lemma* says that on an open set $U \in M$ diffeomorphic to \mathbb{R}^N , all closed forms (cycles) of degree $p \geq 1$ are exact (boundaries). That is, the Betti numbers satisfy $b^p = 0$ (resp. $b = 0$), for $p = 1, \dots, n$.

The *de Rham Theorem* states the following. The map $\Phi: H_n \times H^n \rightarrow \mathbb{R}$ given by $([C], [\omega]) \rightarrow \langle C, \omega \rangle$ for $C \in Z_n, \omega \in Z^n$ is a bilinear nondegenerate map which establishes the duality of the groups (vector spaces) H_n and H^n and the equality $b_n = b^n$.

Configuration Manifold Reduction and its Euler Characteristic

Recall (see subsection (3.8.4.3) above), that for the purpose of high-level control, the rotational biodynamical configuration manifold M (Figure 3.6), could be first, reduced to an n -torus, and second, transformed into an n -cube ‘hyper-joystick’, using the following topological techniques (see [Ivancevic and Pearce (2001b); Ivancevic (2002); Ivancevic (2005)]).

Let S^1 denote the constrained unit circle in the complex plane, which is an Abelian Lie group. Firstly, we propose two reduction homeomorphisms, using the noncommutative semidirect product ‘ \triangleright ’ of the constrained $SO(2)$ -groups:

$$SO(3) \gtrsim SO(2) \triangleright SO(2) \triangleright SO(2), \quad \text{and} \quad SO(2) \approx S^1.$$

Next, let I^n be the unit cube $[0, 1]^n$ in \mathbb{R}^n and ‘ \sim ’ an equivalence relation on \mathbb{R}^n get by ‘gluing’ together the opposite sides of I^n , preserving their orientation. Therefore, the manifold M can be represented as the quotient space of \mathbb{R}^n by the space of the integral lattice points in \mathbb{R}^n , that is an oriented and constrained n D torus T^n :

$$\mathbb{R}^n / Z^n = I^n / \sim \approx \prod_{i=1}^n S_i^1 \equiv \{(q^i, i = 1, \dots, N) : \text{mod } 2\pi\} = T^n. \quad (3.221)$$

Now, using the *de Rham Theorem* and the *homotopy Axiom* for the de Rham cohomologies, we can calculate the *Euler–Poincaré characteristics*

for T^n as well as for its two bundles, TT^n and T^*T^n , as (see [Ivancevic (2002); Ivancevic (2005)])

$$\chi(T^n, TT^n) = \sum_{p=1}^n (-1)^p b_p, \quad \text{where } b_p \text{ are the Betti numbers defined as}$$

$$b^0 = 1, \quad b^1 = n, \dots, \quad b^p = \binom{n}{p}, \dots, \quad b^{n-1} = n, \quad b^n = 1, \quad (p = 0, \dots, n).$$

3.13.5.2 Morse Theory in Biodynamics

Morse Geometry of a Biodynamical Manifold

Recall that on any smooth manifold M there exist many Riemannian metrics g . Each of these metrics is *locally defined* in a particular point $q \in M$ as a symmetric $(0, 2)$ tensor-field such that $g|_q : T_q M \times T_q M \rightarrow \mathbb{R}$ is a positively defined inner product for each point $q \in M$. In an open local chart $U \in M$ containing the point q , this metric is given as $g|_q \mapsto g_{ij}(q) dq^i dq^j$. With each metric $g|_q$ there is associated a *local geodesic* on M .

Now, two main *global geodesics problems* on the biodynamical configuration manifold M with the Riemannian metrics g , can be formulated as follows (compare with subsection 3.10.5.1 above):

- (1) *Is there a minimal geodesic $\gamma_0(t)$ between two points A and B on M ? In other words, does an arc of geodesic $\gamma_0(t)$ with extremities A, B actually have minimum length among all rectifiable curves $\gamma(t) = (q^i(t), p_i(t))$ joining A and B ?*
- (2) *How many geodesic arcs are there joining two points A and B on M ?*

Locally these problems have a *complete answer*: each point of the biodynamics manifold M has an open neighborhood V such that for any two distinct points A, B of V there is exactly *one* arc of a geodesic *contained* in V and joining A and B , and it is the *unique minimal geodesic* between A and B .

Recall (see subsection (3.10.5.1) above), that seven decades ago, Morse considered the set $\Omega = \Omega(M; A, B)$ of *piecewise smooth paths* on a Riemannian manifold M having fixed extremities A, B , defined as continuous maps $\gamma : [0, 1] \rightarrow M$ such that $\gamma(0) = A$, $\gamma(1) = B$, and there were a finite number of points

$$t_0 = 0 < t_1 < t_2 < \dots < t_{m-1} < t_m = 1, \quad (3.222)$$

such that in every *closed interval* $[t_i, t_{i+1}]$, γ was a C^∞ -function. The parametrization was always chosen such that for $t_j \leq t \leq t_{j+1}$,

$$t - t_j = \frac{t_{j+1} - t_j}{l_j} \int_{t_j}^t \left\| \frac{d\gamma}{du} \right\| du, \quad \text{with} \quad l_j = \int_{t_j}^{t_{j+1}} \left\| \frac{d\gamma}{du} \right\| du. \quad (3.223)$$

In other words, $t - t_j$ was proportional to the length of the image of $[t_j, t]$ by γ . Then

$$L(\gamma) = \sum_{j=0}^m l_j,$$

the length of γ , was a function of γ in Ω . A minimal arc from A to B should be a path γ for which $L(\gamma)$ is *minimum* in Ω , and a geodesic arc from A to B should be a path that is a ‘critical point’ for the function L . This at first has no meaning, since Ω is not a differential manifold; the whole of Morse’s theory consists in showing that it is possible to substitute for Ω genuine differential manifolds to which his results on critical points can be applied ([Morse (1934)]).

To study the geodesics joining two points A, B it is convenient, instead of working with the length $L(\gamma)$, to work with the *energy of a path* $\gamma : [A, B] \rightarrow M$, defined by ([Dieudonne (1988)])

$$E_A^B(\gamma) = \int_A^B \left\| \frac{d\gamma}{du} \right\|^2 du. \quad (3.224)$$

With the chosen parametrization (3.223), $E(\gamma) = (B - A)L(\gamma)^2$, and the extremals of E are again the geodesics, but the computations are easier with E .

Morse theory can be divided into several steps (see [Milnor (1963)]).

Step 1 is essentially a presentation of the classical Lagrangian method that brings to light the analogy with the critical points of a C^∞ -function on M . No topology is put on Ω ; a *variation* of a path $\gamma \in \Omega$ is a continuous map α into M , defined in a product $] - \varepsilon, \varepsilon[\times [0, 1]$ with the following properties:

- (1) $\alpha(0, t) = \gamma(t)$;
- (2) $\alpha(u, 0) = A$, $\alpha(u, 1) = B$ for $-\varepsilon < u < \varepsilon$; and
- (3) There is a decomposition (3.222) such that α is C^∞ in each set

$$] - \varepsilon, \varepsilon[\times [t_i, t_{i+1}].$$

A *variation vector-field* $t \mapsto W(t)$ is associated to each variation α , where $W(t)$ is a tangent vector in the tangent space $T_{\gamma(t)}M$ to M , defined by

$$W(t) = \partial_u \alpha(0, t). \quad (3.225)$$

It is a continuous map of $[0, 1]$ into the tangent bundle TM , *smooth* in each interval $[t_i, t_{i+1}]$. These maps are the substitute for the *tangent vectors* at the *point* γ ; they form an infinite-dimensional vector space written $T\Omega(\gamma)$.

More generally the interval $] - \varepsilon, \varepsilon[$ can be replaced in the definition of a variation by a neighborhood of 0 in some \mathbb{R}^n , defining an *n-parameter variation*.

A *critical path* $\gamma_0 \in \Omega$ for a function $F : \Omega \rightarrow \mathbb{R}$ is defined by the condition that for every variation α of γ_0 the function

$$u \mapsto F(\alpha(u, \cdot))$$

is derivable for $u = 0$ and its derivative is 0.

Step 2 is a modern presentation of the formulas of Riemannian geometry, giving the *first variation* and *second variation* of the energy (3.224) of a path $\gamma_0 \in \Omega$, which form the basis of Jacobi results.

First consider an arbitrary path $\omega_0 \in \Omega$, its *velocity* $\dot{\omega}(t) = d\omega/dt$, and its *acceleration* in the Riemannian sense

$$\bar{\ddot{\omega}}(t) = \nabla_t \dot{\omega}(t),$$

where ∇_t denotes the Bianchi covariant derivative. They belong to $T_{\omega(t)}M$ for each $t \in [0, 1]$, are defined and continuous in each interval $[t_i, t_{i+1}]$ in which ω is smooth, and have limits at both extremities. Now let α be a variation of ω and $t \mapsto W(t)$ be the corresponding variation vector-field (3.225). The *first variation formula* gives the first derivative

$$\frac{1}{2} \frac{d}{du} E(\alpha(u, \cdot))|_{u=0} = - \sum_i (W(t_i) | \dot{\omega}(t_i+) - \dot{\omega}(t_i-)) - \int_0^1 (W(t) | \bar{\ddot{\omega}}(t)) dt,$$

where $(x|y)$ denotes the scalar product of two vectors in a tangent space. It follows from this formula that $\gamma_0 \in \Omega$ is a critical path for E iff γ is a *geodesic*.

Next, fix such a geodesic γ and consider a two-parameter variation:

$$\alpha : U \times [0, 1] \rightarrow M,$$

where U is a neighborhood of 0 in \mathbb{R}^2 , so that

$$\alpha(0, 0, t) = \gamma(t), \quad \partial_{u_1}\alpha(0, 0, t) = W_1(t), \quad \partial_{u_2}\alpha(0, 0, t) = W_2(t),$$

in which W_1 and W_2 are in $T\Omega(\gamma)$. The *second variation formula* gives the mixed second derivative

$$\begin{aligned} \frac{1}{2} \frac{\partial^2}{\partial u_1 \partial u_2} E(\alpha(u_1, u_2, \cdot))|_{(0,0)} &= - \sum_i (W_2(t_i) |\nabla_t W_1(t_i+) - \nabla_t W_1(t_i-)) \\ &\quad - \int_0^1 (W_2(t) |\nabla_t^2 W_1(t) + R(V(t) \wedge W_1(t)) \cdot V(t)) dt, \end{aligned} \quad (3.226)$$

where $Z \mapsto R(X \wedge Y) \cdot Z$ is the curvature of the Levi-Civita connection. The l.h.s of (3.226) is thus a *bilinear symmetric form*

$$(W_1, W_2) \mapsto E_{**}(W_1, W_2)$$

on the product $T\Omega(\gamma) \times T\Omega(\gamma)$. For a one-parameter variation α

$$E_{**}(W, W) = \frac{1}{2} \frac{d^2}{du^2} E(\alpha(u, \cdot))|_{u=0},$$

from which it follows that if γ is a *minimal* geodesic in Ω , $E_{**}(W, W) \geq 0$ in $T\Omega(\gamma)$. As usual, we shall speak of E_{**} indifferently as a symmetric bilinear form or as a quadratic form $W \mapsto E_{**}(W, W)$.

Formula (3.226) naturally leads to the junction with Jacobi work (see [Dieudonne (1988)]): consider the smooth vector-fields $t \mapsto J(t)$ along $\gamma \in M$, satisfying the equation

$$\nabla_t^2 J(t) + R(V(t) \wedge J(t)) \cdot V(t) = 0 \quad \text{for } 0 \leq t \leq 1. \quad (3.227)$$

With respect to a frame along γ moving by parallel translation on M this relation is equivalent to a system of n linear homogeneous ODEs of order 2 with C^∞ -coefficients; the solutions J of (3.227) are called the *Jacobi fields* along γ and form a vector space of dimension $2n$. If for a value $a \in]0, 1]$ of the parameter t there exists a Jacobi field along γ that is not identically 0 but *vanishes* for $t = 0$ and $t = a$, then the points $A = \gamma(0)$ and $r = \gamma(a)$ are conjugate along γ with a *multiplicity* equal to the dimension of the vector space of Jacobi fields vanishing for $t = 0$ and $t = a$.

Jacobi fields on the biodynamical configuration manifold M may also be defined as variation vector-fields for *geodesic variations* of the path $\gamma \in M$: they are C^∞ -maps

$$\alpha :] - \varepsilon, \varepsilon[\times [0, 1] \rightarrow M,$$

such that for any $u \in]-\varepsilon, \varepsilon[$, $t \mapsto \alpha(u, t)$ is a geodesic and $\alpha(0, t) = \gamma(t)$.

It can be proved that the Jacobi fields along $\gamma \in M$ that vanish at A and B (hence belong to $T\Omega(\gamma)$) are exactly the vector-fields $J \in T\Omega(\gamma)$ such that

$$E_{**}(J, W) = 0$$

for every $W \in T\Omega(\gamma)$. Although $T\Omega(\gamma)$ is infinite-dimensional, the form E_{**} is again called *degenerate* if the vector space of the Jacobi fields vanishing at A and B is not reduced to 0 and the dimension of that vector space is called the *nullity* of E_{**} . Therefore, E_{**} is thus degenerate iff A and B are conjugate along γ and the nullity of E_{**} is the multiplicity of B .

Step 3 is the beginning of Morse's contributions (see [Milnor (1963)]). He first considered a *fixed* geodesic $\gamma : [0, 1] \rightarrow M$ with extremities $A = \gamma(0)$, $B = \gamma(1)$ and the bilinear symmetric form $E_{**} : T\Omega(\gamma) \times T\Omega(\gamma) \rightarrow \mathbb{R}$. By analogy with the finite-dimensional quadratic form, the *index* of E_{**} is defined as the maximum dimension of a vector subspace of $T\Omega(\gamma)$ in which E_{**} is *strictly negative* (i.e., nondegenerate and taking values $E_{**}(W, W) < 0$ except for $W = 0$). Morse's central result gives the value of the index of E_{**} and is known as the *index Theorem*.

Suppose a subdivision (3.222) is chosen such that each arc $\gamma([t_{i-1}, t_i])$ is contained in an open set $U_i \subset M$ such that any two points of U_i are joined by a unique geodesic arc contained in U_i that is *minimal*; $\gamma([t_{i-1}, t_i])$ is such an arc. In the infinite-dimensional vector space $T\Omega(\gamma)$, consider the two vector subspaces:

- (1) $T\Omega(\gamma; t_0, t_1, \dots, t_m)$ consisting of all continuous vector-fields $t \mapsto W(t)$ along γ , vanishing for $t = 0$ and $t = 1$, such that each restriction $W|_{[t_{i-1}, t_i]}$ is a *Jacobi field* (hence smooth) along $\gamma([t_{i-1}, t_i])$; that subspace is finite-dimensional;
- (2) T' consisting of the vector-fields $t \mapsto W(t)$ along γ , such that $W(t_0) = 0$, $W(t_1) = 0, \dots, W(t_m) = 0$.

$T\Omega(\gamma)$ is then the *direct sum* $T\Omega(\gamma; t_0, t_1, \dots, t_m) \oplus T'$; these two subspaces are orthogonal for the bilinear form E_{**} , and E_{**} is *strictly positive* in T' , so that the index of E_{**} is equal to the index of its restriction to the subspace $T\Omega(\gamma; t_0, t_1, \dots, t_m)$.

To calculate the nullity and index of E_{**} , due to this decomposition, apply their definitions either to vector subspaces of $T\Omega(\gamma)$ or to vector subspaces of $T\Omega(\gamma; t_0, t_1, \dots, t_m)$. The computation of the index of E_{**} is

done by considering the geodesic arc $\gamma_\tau : [0, \tau] \rightarrow M$, the restriction of γ to $[0, \tau]$, and its energy

$$E(\gamma_\tau) = \tau \int_0^\tau \left\| \frac{d\gamma}{du} \right\|^2 du.$$

E_τ^τ is the corresponding quadratic form on $T\Omega(\gamma_\tau)$, and $\lambda(\tau)$ is its index; one studies the variation of $\lambda(\tau)$ when τ varies from 0 to 1, and $\lambda(1)$ is the index of E_{**} .

The index Theorem says: *the index of E_{**} is the sum of the multiplicities of the points conjugate to A along B and distinct from B .*

We have seen that the dimension of $T\Omega(\gamma; t_0, t_1, \dots, t_m)$ is finite; it follows that the index of E_{**} is always *finite*, and therefore the number of points conjugate to A along γ is also *finite*.

Step 4 of Morse theory introduces a *topology* on the set $\Omega = \Omega(M; A, B)$. On the biodynamical configuration manifold M the usual topology can be defined by a *distance* $\rho(A, B)$, the g.l.b. of the lengths of all piecewise smooth paths joining A and B . For any pair of paths ω_1, ω_2 in $\Omega(M; A, B)$, consider the function $d(\omega_1, \omega_2) \in M$

$$d(\omega_1, \omega_2) = \sup_{0 \leq t \leq 1} \rho(\omega_1(t), \omega_2(t)) + \sqrt{\int_0^1 (\dot{s}_1 - \dot{s}_2)^2 dt},$$

where $s_1(t)$ (resp. $s_2(t)$) is the length of the path $\tau \mapsto \omega_1(\tau)$ (resp. $\tau \mapsto \omega_2(\tau)$) defined in $[0, t]$. This distance on Ω such that the function $\omega \mapsto E_A^B(\omega)$ is *continuous* for that distance.

Morse Homology of a Biodynamical Manifold

Morse Functions and Boundary Operators. Let $f : M \rightarrow \mathbb{R}$ represents a C^∞ -function on the biodynamical configuration manifold M . Recall that $z = (q, p) \in M$ is the *critical point* of f if $df(z) \equiv df[(q, p)] = 0$. In local coordinates $(x^1, \dots, x^n) = (q^1, \dots, q^n, p_1, \dots, p_n)$ in a neighborhood of z , this means $\frac{\partial f}{\partial x^i}(z) = 0$ for $i = 1, \dots, n$. The Hessian of f at a critical point z defines a symmetric bilinear form $\nabla df(z) = d^2f(z)$ on T_zM , in local coordinates (x^1, \dots, x^n) represented by the matrix $\left(\frac{\partial^2 f}{\partial x^i \partial x^j} \right)$. Index and nullity of this matrix are called index and nullity of the critical point z of f .

Now, we assume that all critical points z_1, \dots, z_n of $f \in M$ are nondegenerate in the sense that the Hessians $d^2f(z_i)$, $i = 1, \dots, m$, have maximal rank. Let z be such a critical point of f of *Morse index* s ($=$ number

of negative eigenvalues of $d^2f(z_i)$, counted with multiplicity). The eigenvectors corresponding to these negative eigenvalues then span a subspace $V_z \subset T_zM$ of dimension s . We choose an orthonormal basis e_1, \dots, e_s of V_z w.r.t. the Riemannian metric g on M (induced by the system's kinetic energy), with dual basis dx^1, \dots, dx^s . This basis then defines an orientation of V_z which we may also represent by the s -form $dx^1 \wedge \dots \wedge dx^s$. We now let z' be another critical point of f , of Morse index $s-1$. We consider paths $\gamma(t)$ of the steepest descent of f from z to z' , i.e., integral curves of the vector-field $-\nabla f(\gamma)$. Thus $\gamma(t)$ defines the *gradient flow* of f

$$\dot{\gamma}(t) = -\nabla f(\gamma(t)), \quad \text{with} \quad \begin{cases} \lim_{t \rightarrow -\infty} \gamma(t) = z, \\ \lim_{t \rightarrow \infty} \gamma(t) = z' \end{cases}. \quad (3.228)$$

A path $\gamma(t)$ obviously depends on the Riemannian metric g on M as

$$\nabla f = g^{ij} \partial_{x^i} f \partial_{x^j} f.$$

From [Smale (1960); Smale (1967)] it follows that for a generic metric g , the Hessian $\nabla df(y)$ has only nondegenerate eigenvalues. Having a metric g induced by the system's kinetic energy, we let $\tilde{V}_y \subset T_yM$ be the space spanned by the eigenvectors corresponding to the $s-1$ lowest eigenvalues. Since z' has Morse index $s-1$, $\nabla df(z') = d^2f(z')$ has precisely $s-1$ negative eigenvalues. Therefore, $\tilde{V}_{z'} \equiv \lim_{t \rightarrow \infty} \tilde{V}_{\gamma(t)} = V_{z'}$, while the unit tangent vector of γ at z' , i.e., $\lim_{t \rightarrow \infty} \frac{\dot{\gamma}(t)}{\|\dot{\gamma}(t)\|}$, lies in the space of directions corresponding to positive eigenvalues and is thus orthogonal to $V_{z'}$. Likewise, the unit tangent vector v_z of γ at z , while contained in V_z , is orthogonal to \tilde{V}_z , because it corresponds to the largest one among the s negative eigenvalues of $d^2f(z)$. Taking the interior product $i(v_z) dx^1 \wedge \dots \wedge dx^s$ defines an orientation of \tilde{V}_z . Since \tilde{V}_y depends smoothly on y , we may transport the orientation of \tilde{V}_z to $\tilde{V}_{z'}$ along γ . We then define $n_\gamma = +1$ or -1 , depending on whether this orientation of $\tilde{V}_{z'}$ coincides with the chosen orientation of $V_{z'}$ or not, and further define $n(z, z') = \sum_\gamma n_\gamma$, where the sum is taken over all such paths γ of the steepest descent from p to p' .

Now, let M^s be the *set of critical points* of f of *Morse index* s , and let H_f^s be the vector space over \mathbb{R} spanned by the elements of M^s . We define a *boundary operator*

$$\begin{aligned} \delta : H_f^{s-1} &\rightarrow H_f^s, & \text{by putting, for } z' \in M^{s-1}, \\ \delta(z') &= \sum_{n \in M^s} n(z', z) z, & \text{and extending } \delta \text{ by linearity.} \end{aligned}$$

This operator satisfies $\delta^2 = 0$ and therefore defines a cohomology theory. Using *Conley's continuation principle*, Floer [Floer (1988)] showed that the resulting cohomology theories resulting from different choices of f are canonically isomorphic.

In his QFT-based rewriting the Morse topology, Ed Witten [Witten (1982)] considered also the operators:

$$\begin{aligned} d_t &= e^{-tf} de^{tf}, & \text{their adjoints: } d_t^* &= e^{tf} de^{-tf}, \\ & \text{as well as their Laplacian: } \Delta_t &= d_t d_t^* + d_t^* d_t. \end{aligned}$$

For $t = 0$, Δ_0 is the standard *Hodge-de Rham Laplacian*, whereas for $t \rightarrow \infty$, one has the following expansion

$$\Delta_t = dd^* + d^*d + t^2 \|df\|^2 + t \sum_{k,j} \frac{\partial^2 h}{\partial x^k \partial x^j} [i \partial_{x^k}, dx^j],$$

where $(\partial_{x^k})_{k=1,\dots,n}$ is an orthonormal frame at the point under consideration. This becomes very large for $t \rightarrow \infty$, except at the critical points of f , i.e., where $df = 0$. Therefore, the eigenvalues of Δ_t will concentrate near the critical points of f for $t \rightarrow \infty$, and we get an *interpolation* between de Rham cohomology and Morse cohomology.

Morse Homology on M . Now, following [Milinkovic (1999); Ivancevic and Pearce (2006)], for any Morse function f on the configuration manifold M we denote by $\text{Crit}_p(f)$ the set of its critical points of index p and define $C_p(f)$ as a free Abelian group generated by $\text{Crit}_p(f)$. Consider the gradient flow generated by (3.228). Denote by $\mathcal{M}_{f,g}(M)$ the set of all $\gamma : \mathbb{R} \rightarrow M$ satisfying (3.228) such that

$$\int_{-\infty}^{+\infty} \left| \frac{d\gamma}{dt} \right|^2 dt < \infty.$$

The spaces

$$\mathcal{M}_{f,g}(x^-, x^+) = \{\gamma \in \mathcal{M}_{f,g}(M) \mid \gamma(t) \rightarrow x^\pm \text{ as } t \rightarrow \pm\infty\}$$

are smooth manifolds of dimension $m(x^+) - m(x^-)$, where $m(x)$ denotes the Morse index of a critical point x . Note that

$$\mathcal{M}_{f,g}(x, y) \cong W_g^u(x) \cap W_g^s(y),$$

where $W_g^s(y)$ and $W_g^u(x)$ are the stable and unstable manifolds of the gradient flow (3.228). For generic g the intersection above is transverse (Morse–Smale condition). The group \mathbb{R} acts on $\mathcal{M}_{f,g}(x, y)$ by $\gamma \mapsto \gamma(\cdot + t)$. We denote

$$\widehat{\mathcal{M}}_{f,g}(x, y) = \mathcal{M}_{f,g}(x, y)/\mathbb{R}.$$

The manifolds $\widehat{\mathcal{M}}_{f,g}(x, y)$ can be given a coherent orientation σ (see [Schwarz (1993)]).

Now, we can define the boundary operator, as

$$\partial : C_p(f) \rightarrow C_{p-1}(f), \quad \partial x = \sum_{y \in \text{Crit}_{p-1}(f)} n(x, y)y,$$

where $n(x, y)$ is the number of points in 0D manifold $\widehat{\mathcal{M}}_{f,g}(x, y)$ counted with the sign with respect to the orientation σ . The proof of $\partial \circ \partial = 0$ is based on gluing and cobordism arguments [Schwarz (1993)]. Now Morse homology groups are defined by

$$H_p^{\text{Morse}}(f) = \text{Ker}(\partial)/\text{Im}(\partial).$$

For generic choices of Morse functions f_1 and f_2 the groups $H_p(f_1)$ and $H_p(f_2)$ are isomorphic. Furthermore, they are isomorphic to the singular homology group of M , i.e.,

$$H_p^{\text{Morse}}(f) \cong H_p^{\text{sing}}(M),$$

for generic f [Milnor (1965)].

The construction of isomorphism is given (see [Milinkovic (1999); Ivancevic and Pearce (2006)]) as

$$h_{\alpha\beta} : H_p(f^\alpha) \rightarrow H_p(f^\beta), \quad (3.229)$$

for generic Morse functions f^α, f^β . Consider the ‘connecting trajectories’, i.e., the solutions of non-autonomous equation

$$\dot{\gamma} = -\nabla f_t^{\alpha\beta}, \quad (3.230)$$

where $f_t^{\alpha\beta}$ is a homotopy connecting f^α and f^β such that for some $R > 0$

$$f_t^{\alpha\beta} \equiv \begin{cases} f^\alpha & \text{for } t \leq -R \\ f^\beta & \text{for } t \geq R \end{cases}.$$

For $x^\alpha \in \text{Crit}_p(f^\alpha)$ and $x^\beta \in \text{Crit}_p(f^\beta)$ denote

$$\mathcal{M}_{f^{\alpha\beta},g}(x^\alpha, x^\beta) = \{\gamma : \gamma \text{ satisfies (3.230) and } \lim_{t \rightarrow -\infty} \gamma = x^\alpha, \lim_{t \rightarrow \infty} \gamma = x^\beta\}.$$

As before, $\mathcal{M}_{f^{\alpha\beta},g}$ is a smooth finite-dimensional manifold. Now, define

$$\begin{aligned} (h_{\alpha\beta})_\# : C_p(f^\alpha) &\rightarrow C_p(f^\beta), & \text{by} \\ (h_{\alpha\beta})_\# x^\alpha &= \sum_{x^\beta \in \text{Crit}_p(f^\beta)} n(x^\alpha, x^\beta) x^\beta, & \text{for } x^\alpha \in \text{Crit}_p(f^\alpha), \end{aligned}$$

where $n(x^\alpha, x^\beta)$ is the algebraic number of points in 0D manifold $\mathcal{M}_{f^{\alpha\beta},g}(x^\alpha, x^\beta)$ counted with the signs defined by the orientation of $\mathcal{M}_{f^{\alpha\beta},g}$. Homomorphisms $(h_{\alpha\beta})_\#$ commute with ∂ and thus define the homomorphisms $h_{\alpha\beta}$ in homology which, in addition, satisfy $h_{\alpha\beta} \circ h_{\beta\gamma} = h_{\alpha\gamma}$.

Now, if we fix a Morse function $f : M \rightarrow \mathbb{R}$ instead of a metric g , we establish the isomorphism (see [Milinkovic (1999); Ivancevic and Pearce (2006)])

$$h_{\alpha\beta} : H_p(g^\alpha, f) \rightarrow H_p(g^\beta, f)$$

between the two Morse homology groups defined by means of two generic metrics g^α and g^β in a similar way, by considering the ‘connecting trajectories’,

$$\dot{\gamma} = -\nabla^{g_t^{\alpha\beta}} f. \quad (3.231)$$

Here $g_t^{\alpha\beta}$ is a homotopy connecting g^α and g^β such that for some $R > 0$

$$g_t^{\alpha\beta} \equiv \begin{cases} g^\alpha & \text{for } t \leq -R, \\ g^\beta & \text{for } t \geq R, \end{cases}$$

and ∇^g is a gradient defined by metric g .

Note that f is decreasing along the trajectories solving autonomous gradient equation (3.228). Therefore, the boundary operator ∂ preserves the downward filtration given by level sets of f . In other words, if we denote

$$\text{Crit}_p^\lambda(f) = \text{Crit}_p(f) \cap f^{-1}((-\infty, \lambda]), \quad \text{and}$$

$$C_p^\lambda(f) = \text{free Abelian group generated by } \text{Crit}_p^\lambda(f),$$

then the boundary operator ∂ restricts to $\partial^\lambda : C_p^\lambda(f) \rightarrow C_{p-1}^\lambda(f)$. Obviously, $\partial^\lambda \circ \partial^\lambda = 0$, thus we can define the relative Morse homology groups

$$H_p^\lambda(f) = \text{Ker}(\partial^\lambda) / \text{Im}(\partial^\lambda).$$

Following the standard algebraic construction, we define (relative) Morse cohomology. We set

$$C_\lambda^p(f) = \text{Hom}(C_p^\lambda(f), \mathbb{Z}), \quad \text{and} \\ \delta^\lambda : C_\lambda^p(f) \rightarrow C_\lambda^{p+1}(f), \quad \langle \delta^\lambda a, x \rangle = \langle a, \partial^\lambda x \rangle$$

and define

$$H_\lambda^p(f) = \text{Ker}(\delta^\lambda) / \text{Im}(\delta^\lambda).$$

Since $\text{Crit}_p(f)$ is finite, we have $H_p^\lambda(f) = H_p(f)$ and $H_\lambda^p(f) = H^p(f)$.

3.13.5.3 Hodge–De Rham Theory in Biodynamics

Hodge Laplacian

A single biodynamical configuration manifold M can be equipped with many different Riemannian metrics g in local coordinates (apart from the one generated by its kinetic energy)

$$g = g_{ij}(u^1, u^2, \dots, u^n) du^i du^j.$$

Beltrami had shown that it is always possible for such a metric to define an operator (depending on the metric) that generalizes the usual Laplacian on \mathbb{R}^n and therefore induces the notion of harmonic functions on the Riemannian manifold [Choquet-Bruhat and DeWitt-Morete (1982)].

Hodge theory was described by H. Weyl as ‘one of the landmarks in the history of mathematics in the 20th Century’. Hodge showed that it was possible to define a notion of *harmonic exterior differential form*: the metric g on M canonically defines a metric on the tangent bundle TM , hence also, by standard multilinear algebra, a metric on any bundle of tensors on M . In particular, let $(\alpha, \beta) \mapsto g_p(\alpha, \beta)$ be the positive nondegenerate symmetric bilinear form defined on the vector space of p -forms on M . As M is orientable, this defines a *duality* between p -forms and $(n-p)$ -forms: to each p -form α is associated a $(n-p)$ -form $*\alpha$, defined by the linear Hodge star operator $*$ (see subsection 3.6.3.7), characterized by the relations

$$\beta \wedge (*\alpha) = g_p(\alpha, \beta) v, \quad **\alpha = (-1)^{p(n-p)}\alpha,$$

for all p -forms α, β , where v is the volume form on the Riemannian manifold M . If d is the exterior derivative, it has a *transposed* (adjoint) operator

for that duality, the codifferential δ , defined as

$$\delta = -(*) \circ d \circ (*),$$

which maps p -forms onto $(p-1)$ -forms, such that

$$\delta\alpha = (-1)^{np+n+1} * d * \alpha.$$

Recall (from subsection 3.6.2 above) that the *Hodge Laplacian*, defined as

$$\Delta = d \circ \delta + \delta \circ d,$$

transforms p -forms into p -forms and generalizes Beltrami's Laplacian (3.47), which is the special case for $p=0$ (up to a sign). This defines *harmonic* (real or complex valued) p -forms as those for which $\Delta\alpha = 0$, or equivalently, $d\alpha = \delta\alpha = 0$.

In other words, let dv be the volume element of the chosen metric g . Then for every p -form α we can define a *norm functional*

$$\|\alpha\| = \int_X (\alpha, *\alpha)_g dv,$$

for which the *Euler-Lagrangian equation* becomes $\Delta\alpha = 0$.

Now, the p th Betti number of M can be defined as

$$b_p = \dim \text{Ker } \Delta_p,$$

so that the Euler-Poincaré characteristics of M is given by

$$\chi(M) = \sum_{p=0}^n (-1)^p b_p = \sum_{p=0}^n (-1)^p \dim \text{Ker } \Delta_p. \quad (3.232)$$

Finally, for any $(p-1)$ -form α , $(p+1)$ -form β , and harmonic p -form γ ($\Delta\gamma = 0$) on the biodynamical configuration manifold M , the celebrated *Hodge-de Rham decomposition* of a p -form ω [Griffiths (1983b); Voisin (2002)] gives

$$\omega = d\alpha + \delta\beta + \gamma.$$

Now, recall from section 3.14, that a large class of symplectic manifolds is given by the *Kähler manifolds*. Let M be a smooth manifold and g a Riemannian metric on M . Let J be a complex structure on M , that is, $J : TM \rightarrow TM$, $J^2 = -\text{Id}$, and J is g -orthogonal. M is called a *Kähler manifold* if $\nabla J = 0$, where ∇ is the *Levi-Civita connection* of g

and J is regarded as a $(1,1)$ tensor-field. Define a 2-form ω on M by $\omega(X, Y) = g(JX, Y)$, for each vector-field X, Y on M . Then (M, ω) is a symplectic manifold.

Hodge theory takes place on the cohomology of the compact orientable configuration manifold M and reflects the subtle interplay of the following basic additional linear structures one can impose on M :

- Symplectic structure $\omega \in \Gamma_{C^\infty}(M, \Lambda^2 T_M^\vee)$, where ω is nondegenerate, $d\omega = 0$.
- Riemannian structure $g \in \Gamma_{C^\infty}(M, S^2 T_M^\vee)$, where g is positive definite.
- Complex structure $J \in \Gamma_{C^\infty}(M, \text{End}(T_M))$, where $J^2 = -\text{id}$, and J is integrable.

The data (M, ω, g, J) satisfy the *Kähler condition* if ω, g and J are *compatible* in the sense that

$$\omega(\bullet, J(\bullet)) = g(\bullet, \bullet),$$

where \bullet is the strong compatibility condition allowing the comparison of different cohomology theories.

Recall that the de Rham cohomology of (M, J) is defined as

$$H_{DR}^k(M) = \frac{\text{Ker} \left(\Omega^k(M) \xrightarrow{d} \Omega^{k+1}(M) \right)}{\text{Im} \left(\Omega^{k-1}(M) \xrightarrow{d} \Omega^k(M) \right)}.$$

de Rham cohomology classes are represented by harmonic (natural) differential forms.

Let (M, g) be a compact oriented (real or complex) Riemannian manifold. Let dv be the volume element of g . Then for every k -form α we can define

$$\|\alpha\| = \int_M (\alpha, \bar{\alpha})_g dv.$$

The *Euler-Lagrangian equation* for the norm functional turns out to be $d\alpha = \delta\alpha = 0$. A k -form $\alpha \in \Omega^k(M)$ is called *harmonic* if it satisfies one of the following equivalent conditions:

- α is *closed* and $\|\alpha\| \leq \|\alpha + d\beta\|$ for all $\beta \in \Omega^{k-1}(M)$.
- $d\alpha = \delta\alpha = 0$.
- $\Delta\alpha = 0$, where $\Delta = d\delta + \delta d$ is the *Hodge Laplacian*.

Hodge–Weyl Theorem [Griffiths (1983b); Voisin (2002)] states that *every de Rham cohomology class has a unique harmonic representative*.

Heat Kernel and Thermodynamics on M

Besides pure mechanical consideration of biodynamical system, there is another biophysical point of view – thermodynamical, compatible with the human motion [Hill (1938)]. Namely, the *heat equation* on the biodynamical configuration manifold M ,

$$\partial_t a(t) = \Delta a(t), \quad \text{with initial condition } a(0) = \alpha,$$

has a unique solution for every $t \in [0, \infty)$ and every p -form α on M . If we think of α as an *initial temperature distribution* on M then as the configuration manifold cools down, according to the classical heat equation, the temperature should approach a *steady state* which should be *harmonic* [Davies (1989)].

To prove this, we define a *stationary* and hence *harmonic* operator $H(\alpha) = \lim_{t \rightarrow \infty} a(t)$. Also, a map $\alpha \rightarrow G(\alpha)$ with

$$G(\alpha) = \int_0^\infty a(t) dt$$

is *orthogonal to the space of harmonic forms* and satisfies

$$\Delta G(\alpha) = \int_0^\infty \Delta a(t) dt = - \int_0^\infty \partial_t a(t) dt = \alpha - H(\alpha).$$

Here, the map $\alpha \rightarrow H(\alpha)$ is called *harmonic projection* and the map $\alpha \rightarrow G(\alpha)$ is called *Green's operator*.

In particular, for each p -form α we get a *unique decomposition*

$$\alpha = H(\alpha) + \Delta G(\alpha).$$

This proves the existence of a *harmonic representative* in every de Rham cohomology class, as follows.

Let $\alpha \in \Omega^p(M)$ be a closed form. Then

$$\alpha = H(\alpha) + dd^*G(\alpha) + d^*dG(\alpha).$$

But the three terms in this sum are orthogonal and so

$$\|d^*dG(\alpha)\| = \langle d^*dG(\alpha), \alpha \rangle = \langle dG(\alpha), d\alpha \rangle = 0,$$

since α is closed. Thus $H(\alpha)$ is *cohomologous* to α .

This thermal reflection on the biodynamics topology complies with the basic biophysics of human muscles (see [Hill (1938)]).

3.13.5.4 Lagrangian–Hamiltonian Duality in Biodynamics

The present section uncovers the underlying dual geometro–topological structure beneath the general biodynamics. It presents a parallel development of Hamiltonian and Lagrangian formulations of biodynamics (see [Ivancevic and Snoswell (2001); Ivancevic (2002); Ivancevic and Pearce (2001b); Ivancevic and Pearce (2001b); Ivancevic (2005)]), proves both differential–geometrical and algebraic–topo–logical dualities between these two formulations, and finally establishes a *unique functorial relation* between biodynamics geometry and biodynamics topology.

Lagrangian formulation of biodynamics is performed on the *tangent bundle* TM , while Hamiltonian formulation is performed on the *cotangent bundle* T^*M . Both *Riemannian* and *symplectic* geometry are used. The geometrical duality (see [Kolar *et al.* (1993); Choquet-Bruhat and DeWitt-Morete (1982)]) of Lie groups and algebras between these two biodynamics formulations is proved as an existence of natural equivalence between Lie and canonical functors. The topological duality (see [Dodson and Parker (1997)]) between these two biodynamics formulations is proved as an existence of natural equivalence between Lagrangian and Hamiltonian functors in both *homology* and *cohomology* categories. In the case of reduced configuration manifold, the Betti numbers and Euler–Poincaré characteristic are given.

Geometrical Duality Theorem for M

Theorem. There is a geometrical duality between rotational Lagrangian and Hamiltonian biodynamical formulations on M (as given by Figure 3.6). In categorical terms, there is a unique natural geometrical equivalence

$$\text{Dual}_G : \text{Lie} \cong \text{Can}$$

in biodynamics (symbols are described in the next subsection).

Proof. The proof has two parts: Lie–functorial and geometrical.

Lie–Functorial Proof. If we apply the functor **Lie** on the category $\bullet[SO(n)^i]$ (for $n = 2, 3$ and $i = 1, \dots, N$) of rotational Lie groups $SO(n)^i$ (and their homomorphisms) we get the category $\bullet[so(n)_i]$ of corresponding *tangent* Lie algebras $so(n)_i$ (and their homomorphisms). If we further apply the

isomorphic functor **Dual** to the category $\bullet[so(n)_i]$ we get the dual category $\bullet[so(n)_i^*]$ of *cotangent*, or, *canonical* Lie algebras $so(n)_i^*$ (and their homomorphisms). To go directly from $\bullet[SO(n)^i]$ to $\bullet[so(n)_i^*]$ we use the canonical functor **Can**. Therefore, we have a commutative triangle:

$$\begin{array}{ccc}
 & \bullet[SO(n)^i] & \\
 \text{Lie} \swarrow & & \searrow \text{Can} \\
 \bullet[so(n)_i] & \xrightarrow[\cong]{\text{Dual}_A} & \bullet[so(n)_i^*]
 \end{array}$$

\mathcal{LGA}

Applying the functor **Lie** on the biodynamical configuration manifold M , we get the product-tree of the same anthropomorphic structure, but having tangent Lie algebras $so(n)_i$ as vertices, instead of the groups $SO(n)^i$. Again, applying the functor **Can** on M , we get the product-tree of the same anthropomorphic structure, but this time having cotangent Lie algebras $so(n)_i^*$ as vertices. Both the tangent algebras $so(n)_i$ and the cotangent algebras $so(n)_i^*$ contain infinitesimal group generators: angular velocities $\dot{q}^i = \dot{q}^{\phi_i}$ – in the first case, and canonical angular momenta $p_i = p_{\phi_i}$ – in the second case [Ivancevic and Snoswell (2001)]. As Lie group generators, both the angular velocities and the angular momenta satisfy the commutation relations: $[\dot{q}^{\phi_i}, \dot{q}^{\psi_i}] = \epsilon_{\theta}^{\phi\psi} \dot{q}^{\theta_i}$ and $[p_{\phi_i}, p_{\psi_i}] = \epsilon_{\phi\psi}^{\theta} p_{\theta_i}$, respectively, where the structure constants $\epsilon_{\theta}^{\phi\psi}$ and $\epsilon_{\phi\psi}^{\theta}$ constitute the totally antisymmetric third-order tensors.

In this way, the functor $\text{Dual}_G : \text{Lie} \cong \text{Can}$ establishes the unique geometrical duality between kinematics of angular velocities \dot{q}^i (involved in *Lagrangian* formalism on the tangent bundle of M) and kinematics of angular momenta p_i (involved in *Hamiltonian* formalism on the cotangent bundle of M), which is analyzed below. In other words, we have two functors, **Lie** and **Can**, from the *category of Lie groups* (of which $\bullet[SO(n)^i]$ is a subcategory) into the *category of (their) Lie algebras* (of which $\bullet[so(n)_i]$ and $\bullet[so(n)_i^*]$ are subcategories), and a unique natural equivalence between them defined by the functor Dual_G . (As angular momenta p_i are in a bijective correspondence with angular velocities \dot{q}^i , every component of the functor Dual_G is invertible.) ■

Geometrical Proof. Geometrical proof is given along the lines of Riemannian and symplectic geometry of mechanical systems, as follows (see 3.13.1

above, as well as [Marsden and Ratiu (1999); Ivancevic and Snoswell (2001); Ivancevic (2002); Ivancevic and Pearce (2001b); Ivancevic (2005)]). Recall that the Riemannian metric $g = \langle, \rangle$ on the configuration manifold M is a positive-definite quadratic form $g : TM \rightarrow \mathbb{R}$, given in local coordinates $q^i \in U$ (U open in M) as

$$g_{ij} \mapsto g_{ij}(q, m) dq^i dq^j, \quad \text{where}$$

$$g_{ij}(q, m) = m_\mu \delta_{rs} \frac{\partial x^r}{\partial q^i} \frac{\partial x^s}{\partial q^j}$$

is the covariant material metric tensor g , defining a relation between internal and external coordinates and including n segmental masses m_μ . The quantities x^r are external coordinates ($r, s = 1, \dots, 6n$) and $i, j = 1, \dots, N \equiv 6n - h$, where h denotes the number of holonomic constraints.

The *Lagrangian* of the system is a quadratic form $L : TM \rightarrow \mathbb{R}$ dependent on velocity v and such that $L(v) = \frac{1}{2} \langle v, v \rangle$. It is given by

$$L(v) = \frac{1}{2} g_{ij}(q, m) v^i v^j$$

in local coordinates $q^i, v^i = \dot{q}^i \in U_v$ (U_v open in TM). The *Hamiltonian* of the system is a quadratic form $H : T^*M \rightarrow \mathbb{R}$ dependent on momentum p and such that $H(p) = \frac{1}{2} \langle p, p \rangle$. It is given by

$$H(p) = \frac{1}{2} g^{ij}(q, m) p_i p_j$$

in local canonical coordinates $q^i, p_i \in U_p$ (U_p open in T^*M). The inverse (contravariant) metric tensor g^{-1} , is defined as

$$g^{ij}(q, m) = m_\mu \delta_{rs} \frac{\partial q^i}{\partial x^r} \frac{\partial q^j}{\partial x^s}.$$

For any smooth function L on TM , the *fibre derivative*, or *Legendre transformation*, is a diffeomorphism $\mathbb{F}L : TM \rightarrow T^*M$, $\mathbb{F}L(w) \cdot v = \langle w, v \rangle$, from the momentum phase-space manifold to the velocity phase-space manifold associated with the metric $g = \langle, \rangle$. In local coordinates $q^i, v^i = \dot{q}^i \in U_v$ (U_v open in TM), $\mathbb{F}L$ is given by $(q^i, v^i) \mapsto (q^i, p_i)$.

Recall that on the momentum phase-space manifold T^*M exists:

- (i) A unique canonical 1-form θ_H with the property that, for any 1-form β on the configuration manifold M , we have $\beta^* \theta_H = \beta$. In local canonical coordinates $q^i, p_i \in U_p$ (U_p open in T^*M) it is given by $\theta_H = p_i dq^i$.
- (ii) A unique nondegenerate Hamiltonian symplectic 2-form ω_H , which is

closed ($d\omega_H = 0$) and exact ($\omega_H = d\theta_H = dp_i \wedge dq^i$). Each body segment has, in the general $SO(3)$ case, a sub-phase-space manifold $T^*SO(3)$ with

$$\omega_H^{(sub)} = dp_\phi \wedge d\phi + dp_\psi \wedge d\psi + dp_\theta \wedge d\theta.$$

Analogously, on the velocity phase-space manifold TM exists:

- (i) A unique 1-form θ_L , defined by the pull-back $\theta_L = (\mathbb{F}L) * \theta_H$ of θ_H by $\mathbb{F}L$. In local coordinates $q^i, v^i = \dot{q}^i \in U_v$ (U_v open in TM) it is given by $\theta_L = L_{v^i} dq^i$, where $L_{v^i} \equiv \partial L / \partial v^i$.
- (ii) A unique nondegenerate Lagrangian symplectic 2-form ω_L , defined by the pull-back $\omega_L = (\mathbb{F}L) * \omega_H$ of ω_H by $\mathbb{F}L$, which is closed ($d\omega_L = 0$) and exact ($\omega_L = d\theta_L = dL_{v^i} \wedge dq^i$).

Both T^*M and TM are orientable manifolds, admitting the standard volumes given respectively by

$$\Omega_{\omega_H} = \frac{(-1)^{\frac{N(N+1)}{2}}}{N!} \omega_H^N, \quad \text{and} \quad \Omega_{\omega_L} = \frac{(-1)^{\frac{N(N+1)}{2}}}{N!} \omega_L^N,$$

in local coordinates $q^i, p_i \in U_p$ (U_p open in T^*M), resp. $q^i, v^i = \dot{q}^i \in U_v$ (U_v open in TM). They are given by

$$\begin{aligned} \Omega_H &= dq^1 \wedge \cdots \wedge dq^N \wedge dp_1 \wedge \cdots \wedge dp_N, & \text{and} \\ \Omega_L &= dq^1 \wedge \cdots \wedge dq^N \wedge dv^1 \wedge \cdots \wedge dv^N. \end{aligned}$$

On the velocity phase-space manifold TM we can also define the *action* $A : TM \rightarrow \mathbb{R}$ by $A(v) = \mathbb{F}L(v) \cdot v$ and the energy $E = A - L$. In local coordinates $q^i, v^i = \dot{q}^i \in U_v$ (U_v open in TM) we have $A = v^i L_{v^i}$, so $E = v^i L_{v^i} - L$. The Lagrangian vector-field X_L on TM is determined by the condition $i_{X_L} \omega_L = dE$. Classically, it is given by the second-order Lagrangian equations

$$\frac{d}{dt} L_{v^i} = L_{q^i}. \quad (3.233)$$

The Hamiltonian vector-field X_H is defined on the momentum phase-space manifold T^*M by the condition $i_{X_H} \omega = dH$. The condition may be expressed equivalently as $X_H = J \nabla H$, where $J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$.

In local canonical coordinates $q^i, p_i \in U_p$ (U_p open in T^*M) the vector-field X_H is classically given by the first-order Hamiltonian canonical equations

$$\dot{q}^i = \partial_{p_i} H, \quad \dot{p}_i = -\partial_{q^i} H. \quad (3.234)$$

As a Lie group, the configuration manifold M is Hausdorff. Therefore for $x = (q^i, p_i) \in U_p$ (U_p open in T^*M) there exists a unique one-parameter group of diffeomorphisms $\phi_t : T^*M \rightarrow T^*M$ such that $\frac{d}{dt}|_{t=0} \phi_t x = J\nabla H(x)$. This is termed *Hamiltonian phase-flow* and represents the maximal integral curve $t \mapsto (q^i(t), p_i(t))$ of the Hamiltonian vector-field X_H passing through the point x for $t = 0$.

The flow ϕ_t is *symplectic* if ω_H is constant along it (that is, $\phi_t^* \omega_H = \omega_H$) iff its Lie derivative vanishes (that is, $\mathcal{L}_{X_H} \omega_H = 0$). A symplectic flow consists of canonical transformations on T^*M , that is, local diffeomorphisms that leave ω_H invariant. By Liouville Theorem, a symplectic flow ϕ_t preserves the phase volume on T^*M . Also, the total energy $H = E$ of the system is conserved along ϕ_t , that is, $H \circ \phi_t = \phi_t$.

Lagrangian flow can be defined analogously (see [Abraham and Marsden (1978); Marsden and Ratiu (1999)]).

For a Lagrangian (resp. a Hamiltonian) vector-field X_L (resp. X_H) on M , there is a base integral curve $\gamma_0(t) = (q^i(t), v^i(t))$ (resp. $\gamma_0(t) = (q^i(t), p_i(t))$) iff $\gamma_0(t)$ is a geodesic. This is given by the contravariant velocity equation

$$\dot{q}^i = v^i, \quad \dot{v}^i + \Gamma_{jk}^i v^j v^k = 0, \quad (3.235)$$

in the former case, and by the covariant momentum equation

$$\dot{q}^k = g^{ki} p_i, \quad \dot{p}_i + \Gamma_{jk}^i g^{jl} g^{km} p_l p_m = 0, \quad (3.236)$$

in the latter. As before, Γ_{jk}^i denote the Christoffel symbols of an affine connection ∇ in an open chart U on M , defined by the Riemannian metric $g = \langle, \rangle$ as: $\Gamma_{jk}^i = g^{il} \Gamma_{jkl}$, $\Gamma_{jkl} = \frac{1}{2} (\partial_{q^j} g_{kl} + \partial_{q^k} g_{jl} - \partial_{q^l} g_{jk})$.

The l.h.s $\dot{v}^i = \dot{v}^i + \Gamma_{jk}^i v^j v^k$ (resp. $\dot{p}_i = \dot{p}_i + \Gamma_{jk}^i g^{jl} g^{km} p_l p_m$) in the second parts of (3.235) and (3.236) represent the *Bianchi covariant derivative* of the velocity (resp. momentum) with respect to t . Parallel transport on M is defined by $\dot{v}^i = 0$, (resp. $\dot{p}_i = 0$). When this applies, X_L (resp. X_H) is called the *geodesic spray* and its flow the *geodesic flow*.

For the dynamics in the gravitational potential field $V : M \rightarrow \mathbb{R}$, the Lagrangian $L : TM \rightarrow \mathbb{R}$ (resp. the Hamiltonian $H : T^*M \rightarrow \mathbb{R}$) has an extended form

$$\begin{aligned} L(v, q) &= \frac{1}{2} g_{ij} v^i v^j - V(q), \\ (\text{resp. } H(p, q) &= \frac{1}{2} g^{ij} p_i p_j + V(q)). \end{aligned}$$

A Lagrangian vector-field X_L (resp. Hamiltonian vector-field X_H) is still defined by the second-order Lagrangian equations (3.233, 3.235) (resp. first-order Hamiltonian equations (3.234, 3.236)).

The fibre derivative $\mathbb{F}L : TM \rightarrow T^*M$ thus maps Lagrangian equations (3.233, 3.235) into Hamiltonian equations (3.234, 3.236). Clearly there exists a diffeomorphism $\mathbb{F}H : T^*M \rightarrow TM$, such that $\mathbb{F}L = (\mathbb{F}H)^{-1}$. In local canonical coordinates $q^i, p_i \in U_p$ (U_p , open in T^*M) this is given by $(q^i, p_i) \mapsto (q^i, v^i)$ and thus maps Hamiltonian equations (3.234, 3.236) into Lagrangian equations (3.233, 3.235).

A general form of the forced, non-conservative Hamiltonian equations (resp. Lagrangian equations) is given as

$$\begin{aligned} \dot{q}^i &= \frac{\partial H}{\partial p_i}, & \dot{p}_i &= -\frac{\partial H}{\partial q^i} + F_i(t, q^i, p_i), \\ (\text{resp. } \frac{d}{dt} \frac{\partial L}{\partial v^i} - \frac{\partial L}{\partial q^i} &= F_i(t, q^i, v^i)). \end{aligned}$$

Here the $F_i(t, q^i, p_i)$ (resp. $F_i(t, q^i, v^i)$) represent any kind of *covariant forces*, including dissipative and elastic joint forces, as well as actuator drives and control forces, as a function of time, coordinates and momenta. In covariant form we have

$$\begin{aligned} \dot{q}^k &= g^{ki} p_i, & \dot{p}_i + \Gamma_{jk}^i g^{jl} g^{km} p_l p_m &= F_i(t, q^i, p_i), \\ (\text{resp. } \dot{q}^i &= v^i, & \dot{v}^i + \Gamma_{jk}^i v^j v^k &= g^{ij} F_j(t, q^i, v^i)). \quad \blacksquare \end{aligned}$$

This proves the existence of the unique natural geometrical equivalence

$$\text{Dual}_G : \text{Lie} \cong \text{Can}$$

in the rotational biodynamics.

Topological Duality Theorem for M

In this section we want to prove that the general biodynamics can be *equivalently* described in terms of two *topologically dual functors* **Lag** and **Ham**, from **Diff**, the *category of smooth manifolds* (and their smooth maps) of class C^p , into **Bund**, the *category of vector bundles* (and vector-bundle maps) of class C^{p-1} , with $p \geq 1$. **Lag** is physically represented by the second-order Lagrangian formalism on $TM \in \text{Bund}$, while **Ham** is physically represented by the first-order Hamiltonian formalism on $T^*M \in \text{Bund}$.

Theorem. There is a topological duality between Lagrangian and Hamiltonian formalisms on M (as given by Figure 3.6). In categorical terms, there is a unique natural topological equivalence

$$\text{Dual}_T : \text{Lag} \cong \text{Ham}$$

in the general biodynamics.

Proof. The proof has two parts: cohomological and homological.

Cohomological Proof. If $\mathcal{C} = \mathcal{H}^\bullet \mathcal{M}$ (resp. $\mathcal{C} = \mathcal{L}^\bullet \mathcal{M}$) represents the Abelian category of cochains on the momentum phase-space manifold T^*M (resp. the velocity phase-space manifold TM), we have the category $\mathcal{S}^\bullet(\mathcal{H}^\bullet \mathcal{M})$ (resp. $\mathcal{S}^\bullet(\mathcal{L}^\bullet \mathcal{M})$) of generalized cochain complexes A^\bullet in $\mathcal{H}^\bullet \mathcal{M}$ (resp. $\mathcal{L}^\bullet \mathcal{M}$) and if $A' = 0$ for $n < 0$ we have a subcategory $\mathcal{S}_{\mathcal{DR}}^\bullet(\mathcal{H}^\bullet \mathcal{M})$ (resp. $\mathcal{S}_{\mathcal{DR}}^\bullet(\mathcal{L}^\bullet \mathcal{M})$) of de Rham differential complexes in $\mathcal{S}^\bullet(\mathcal{H}^\bullet \mathcal{M})$ (resp. $\mathcal{S}^\bullet(\mathcal{L}^\bullet \mathcal{M})$)

$$\begin{aligned} A_{DR}^\bullet : 0 \rightarrow \Omega^0(T^*M) \xrightarrow{d} \Omega^1(T^*M) \xrightarrow{d} \\ \xrightarrow{d} \Omega^2(T^*M) \xrightarrow{d} \dots \xrightarrow{d} \Omega^N(T^*M) \xrightarrow{d} \dots \\ (\text{resp. } A_{DR}^\bullet : 0 \rightarrow \Omega^0(TM) \xrightarrow{d} \Omega^1(TM) \xrightarrow{d} \Omega^2(TM) \xrightarrow{d} \\ \dots \xrightarrow{d} \Omega^N(TM) \xrightarrow{d} \dots), \end{aligned}$$

where $A' = \Omega^N(T^*M)$ (resp. $A' = \Omega^N(TM)$) is the vector space of all N -forms on T^*M (resp. TM) over \mathbb{R} .

Let $Z^N(T^*M) = \text{Ker}(d)$ (resp. $Z^N(T) = \text{Ker}(d)$) and $B^N(T^*M) = \text{Im}(d)$ (resp. $B^N(TM) = \text{Im}(d)$) denote respectively the real vector spaces of cocycles and coboundaries of degree N . Since $d_{N+1}d_N = d^2 = 0$, it follows that $B^N(T^*M) \subset Z^N(T^*M)$ (resp. $B^N(TM) \subset Z^N(TM)$). The quotient vector space

$$\begin{aligned} H_{DR}^N(T^*M) &= \text{Ker}(d)/\text{Im}(d) = Z^N(T^*M)/B^N(T^*M) \\ (\text{resp. } H_{DR}^N(TM) &= \text{Ker}(d)/\text{Im}(d) = Z^N(TM)/B^N(TM)), \end{aligned}$$

we refer to as the de Rham cohomology group (vector space) of T^*M (resp. TM). The elements of $H_{DR}^N(T^*M)$ (resp. $H_{DR}^N(TM)$) are equivalence sets of cocycles. Two cocycles ω_1 and ω_2 are cohomologous, or belong to the same equivalence set (written $\omega_1 \sim \omega_2$) iff they differ by a coboundary $\omega_1 - \omega_2 = d\theta$. Any form $\omega_H \in \Omega^N(T^*M)$ (resp. $\omega_L \in \Omega^N(TM)$) has a de Rham cohomology class $[\omega_H] \in H_{DR}^N(T^*M)$ (resp. $[\omega_L] \in H_{DR}^N(TM)$).

Hamiltonian symplectic form $\omega_H = dp_i \wedge dq_i$ on T^*M (resp. Lagrangian symplectic form $\omega_L = dL_{v^i} \wedge dq^i$ on TM) is by definition both

a closed 2-form or two-cocycle and an exact 2-form or two-coboundary. Therefore the 2D-de Rham cohomology group of human motion is defined as a quotient vector space

$$\begin{aligned} H_{DR}^2(T^*M) &= Z^2(T^*M)/B^2(T^*M) \\ (\text{resp. } H_{DR}^2(TM) &= Z^2(TM)/B^2(TM)). \end{aligned}$$

As T^*M (resp. TM) is a compact Hamiltonian symplectic (resp. Lagrangian symplectic) manifold of dimension $2N$, it follows that ω_H^N (resp. ω_L^N) is a volume element on T^*M (resp. TM), and the 2ND de Rham's cohomology class $[\omega_H^N] \in H_{DR}^{2N}(T^*M)$ (resp. $[\omega_L^N] \in H_{DR}^{2N}(TM)$) is nonzero. Since $[\omega_H^N] = [\omega_H]^N$ (resp. $[\omega_L^N] = [\omega_L]^N$), then $[\omega_H] \in H_{DR}^2(T^*M)$ (resp. $[\omega_L] \in H_{DR}^2(TM)$) and all of its powers up to the N -th must be zero as well. The existence of such an element is a necessary condition for T^*M (resp. TM) to admit a Hamiltonian symplectic structure ω_H (resp. Lagrangian symplectic structure ω_L).

The de Rham complex A_{DR}^\bullet on T^*M (resp. TM) can be considered as a system of second-order ODEs $d^2\theta_H = 0$, $\theta_H \in \Omega^N(T^*M)$ (resp. $d^2\theta_L = 0$, $\theta_L \in \Omega^N(TM)$) having a solution represented by $Z^N(T^*M)$ (resp. $Z^N(TM)$). In local coordinates q^i , $p_i \in U_p$ (U_p open in T^*M) (resp. q^i , $v^i \in U_v$ (U_v open in TM)) we have $d^2\theta_H = d^2(p_i dq^i) = d(dp_i \wedge dq^i) = 0$, (resp. $d^2\theta_L = d^2(L_{v^i} dq^i) = d(dL_{v^i} \wedge dq^i) = 0$). ■

Homological Proof. If $\mathcal{C} = \mathcal{H}_\bullet\mathcal{M}$, (resp. $\mathcal{C} = \mathcal{L}_\bullet\mathcal{M}$) represents an Abelian category of chains on T^*M (resp. TM), we have a category $\mathcal{S}_\bullet(\mathcal{H}_\bullet\mathcal{M})$ (resp. $\mathcal{S}_\bullet(\mathcal{L}_\bullet\mathcal{M})$) of generalized chain complexes \mathcal{A}_\bullet in $\mathcal{H}_\bullet\mathcal{M}$ (resp. $\mathcal{L}_\bullet\mathcal{M}$), and if $A = 0$ for $n < 0$ we have a subcategory $\mathcal{S}_\bullet^C(\mathcal{H}_\bullet\mathcal{M})$ (resp. $\mathcal{S}_\bullet^C(\mathcal{L}_\bullet\mathcal{M})$) of chain complexes in $\mathcal{H}_\bullet\mathcal{M}$ (resp. $\mathcal{L}_\bullet\mathcal{M}$)

$$\begin{aligned} \mathcal{A}_\bullet : 0 \leftarrow C^0(T^*M) \xleftarrow{\partial} C^1(T^*M) \xleftarrow{\partial} C^2(T^*M) \xleftarrow{\partial} \dots \\ \dots \xleftarrow{\partial} C^n(T^*M) \xleftarrow{\partial} \dots \\ (\text{resp. } \mathcal{A}_\bullet : 0 \leftarrow C^0(TM) \xleftarrow{\partial} C^1(TM) \xleftarrow{\partial} C^2(TM) \xleftarrow{\partial} \dots \\ \dots \xleftarrow{\partial} C^n(TM) \xleftarrow{\partial} \dots). \end{aligned}$$

Here $A_N = C^N(T^*M)$ (resp. $A_N = C^N(TM)$) is the vector space of all finite chains C on T^*M (resp. TM) over \mathbb{R} , and $\partial_N = \partial : C^{N+1}(T^*M) \rightarrow C^N(T^*M)$ (resp. $\partial_N = \partial : C^{N+1}(TM) \rightarrow C^N(TM)$). A finite chain C such that $\partial C = 0$ is an N -cycle. A finite chain C such that $C = \partial B$ is an N -boundary. Let $Z_N(T^*M) = \text{Ker}(\partial)$ (resp. $Z_N(TM) = \text{Ker}(\partial)$) and $B_N(T^*M) = \text{Im}(\partial)$ (resp. $B_N(TM) = \text{Im}(\partial)$) denote respectively

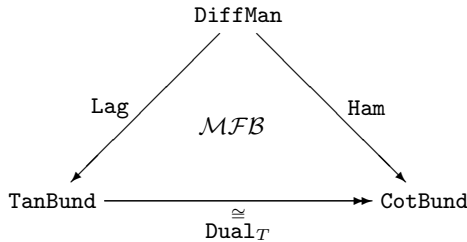
real vector spaces of cycles and boundaries of degree N . Since $\partial_{N-1}\partial_N = \partial^2 = 0$, then $B_N(T^*M) \subset Z_N(T^*M)$ (resp. $B_N(TM) \subset Z_N(TM)$). The quotient vector space

$$\begin{aligned} H_N^C(T^*M) &= Z_N(T^*M)/B_N(T^*M) \\ (\text{resp. } H_N^C(TM) &= Z_N(TM)/B_N(TM)) \end{aligned}$$

represents an *ND biodynamics homology group* (vector space). The elements of $H_N^C(T^*M)$ (resp. $H_N^C(TM)$) are equivalence sets of cycles. Two cycles C_1 and C_2 are homologous, or belong to the same equivalence set (written $C_1 \sim C_2$) iff they differ by a boundary $C_1 - C_2 = \partial B$. The homology class of a finite chain $C \in C^N(T^*M)$ (resp. $C \in C^N(TM)$) is $[C] \in H_N^C(T^*M)$ (resp. $[C] \in H_N^C(TM)$). ■

Lagrangian Versus Hamiltonian Duality

In this way, we have proved a commutativity of a triangle:



which implies the existence of the unique natural topological equivalence

$$\text{Dual}_T : \text{Lag} \cong \text{Ham}$$

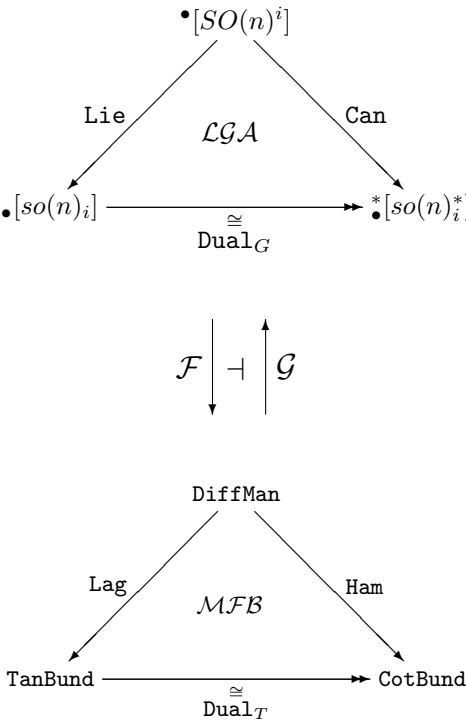
in the rotational biodynamics.

Globally Dual Structure of Rotational Biodynamics

Theorem. Global dual structure of the rotational biodynamics is defined by the unique natural equivalence

$$\text{Dyn} : \text{Dual}_G \cong \text{Dual}_T.$$

Proof. This unique functorial relation, uncovering the natural equivalence between *geometrical* and *topological* structures of biodynamics:



– has been established by parallel development of Lagrangian and Hamiltonian biodynamics formulations, i.e., functors $\text{Lag}(\text{Lie})$ and $\text{Ham}(\text{Can})$. ■

3.14 Complex and Kähler Manifolds and Their Applications

Just as a smooth manifold has enough structure to define the notion of differentiable functions, a *complex manifold* is one with enough structure to define the notion of holomorphic (or, analytic) functions $f : X \rightarrow \mathbb{C}$. Namely, if we demand that the transition functions $\phi_j \circ \phi_i^{-1}$ in the charts U_i on M (see Figure 3.12) satisfy the *Cauchy–Riemann equations*

$$\partial_x u = \partial_y v, \qquad \partial_y u = -\partial_x v,$$

then the analytic properties of f can be studied using its coordinate representative $f \circ \phi_i^{-1}$ with assurance that the conclusions drawn are patch

independent. Introducing local complex coordinates in the charts U_i on M , the ϕ_i can be expressed as maps from U_i to an open set in $\mathbb{C}^{\frac{n}{2}}$, with $\phi_j \circ \phi_i^{-1}$ being a holomorphic map from $\mathbb{C}^{\frac{n}{2}}$ to $\mathbb{C}^{\frac{n}{2}}$. Clearly, n must be even for this to make sense. In local complex coordinates, we recall that a function $h : \mathbb{C}^{\frac{n}{2}} \rightarrow \mathbb{C}^{\frac{n}{2}}$ is *holomorphic* if $h(z^1, \bar{z}^1, \dots, z^{\frac{n}{2}}, \bar{z}^{\frac{n}{2}})$ is actually independent of all the \bar{z}^j .

In a given patch on any even-dimensional manifold, we can always introduce local complex coordinates by, for instance, forming the combinations $z^j = x^j + ix^{\frac{n}{2}+j}$, where the x^j are local real coordinates on M . The real test is whether the transition functions from one patch to another — when expressed in terms of the local complex coordinates — are holomorphic maps. If they are, we say that M is a complex manifold of complex dimension $d = n/2$. The local complex coordinates with holomorphic transition functions give M with a *complex structure* (see [Greene (1996)]).

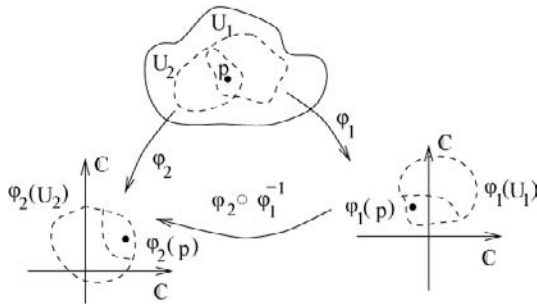


Fig. 3.12 The charts for a complex manifold M have complex coordinates (see text for explanation).

Given a smooth manifold with even real dimension n , it can be a difficult question to determine whether or not a complex structure exists. On the other hand, if some smooth manifold M does admit a complex structure, we are not able to decide whether it is unique, i.e., there may be numerous inequivalent ways of defining complex coordinates on M .

Now, in the same way as a *homeomorphism* defines an equivalence between topological manifolds, and a *diffeomorphism* defines an equivalence between smooth manifolds, a *biholomorphism* defines an equivalence between complex manifolds. If M and N are complex manifolds, we consider them to be equivalent if there is a map $\phi : M \rightarrow N$ which in addition to being a diffeomorphism, is also a holomorphic map. That is, when expressed in terms of the complex structures on M and N respectively, ϕ is

holomorphic. It is not hard to show that this necessarily implies that ϕ^{-1} is holomorphic as well and hence ϕ is known as a biholomorphism. Such a map allows us to identify the complex structures on M and N and hence they are isomorphic as complex manifolds.

These definitions are important because there are pairs of smooth manifolds M and N which are homeomorphic but not diffeomorphic, as well as, there are complex manifolds M and N which are diffeomorphic but not biholomorphic. This means that if one simply ignored the fact that M and N admit local complex coordinates (with holomorphic transition functions), and one only worked in real coordinates, there would be no distinction between M and N . The difference between them only arises from the way in which complex coordinates have been laid down upon them.

Again, recall that a *tangent space* to a manifold M at a point p is the closest flat approximation to M at that point. A convenient basis for the tangent space of M at p consists of the n linearly independent partial derivatives,

$$T_p M : \{\partial_{x^1}|_p, \dots, \partial_{x^n}|_p\}. \quad (3.237)$$

A vector $v \in T_p M$ can then be expressed as $v = v^\alpha \partial_{x^\alpha}|_p$.

Also, a convenient basis for the dual, *cotangent space* $T_p^* M$, is the basis of one-forms, which is dual to (3.237) and usually denoted by

$$T_p^* M : \{dx^1|_p, \dots, dx^n|_p\}, \quad (3.238)$$

where, by definition, $dx^i : T_p M \rightarrow \mathbb{R}$ is a linear map with $dx_p^i(\partial_{x^j}|_p) = \delta_j^i$.

Now, if M is a complex manifold of complex dimension $d = n/2$, there is a notion of the *complexified tangent space* of M , denoted by $T_p M^{\mathbb{C}}$, which is the same as the real tangent space $T_p M$ except that we allow complex coefficients to be used in the vector space manipulations. This is often denoted by writing $T_p M^{\mathbb{C}} = T_p M \otimes \mathbb{C}$. We can still take our basis to be as in (3.237) with an arbitrary vector $v \in T_p M^{\mathbb{C}}$ being expressed as $v = v^\alpha \frac{\partial}{\partial x^\alpha}|_p$, where the v^α can now be complex numbers. In fact, it is convenient to rearrange the basis vectors in (3.237) to more directly reflect the underlying complex structure. Specifically, we take the following linear combinations of basis vectors in (3.237) to be our new basis vectors:

$$\begin{aligned} T_p M^{\mathbb{C}} \quad : \quad & \{(\partial_{x^1} + i\partial_{x^{d+1}})|_p, \dots, \\ & (\partial_{x^d} + i\partial_{x^{2d}})|_p, (\partial_{x^1} - i\partial_{x^{d+1}})|_p, \dots, (\partial_{x^d} - i\partial_{x^{2d}})|_p\}. \end{aligned} \quad (3.239)$$

In terms of complex coordinates we can write the basis (3.239) as

$$T_p M^{\mathbb{C}} : \{\partial_{z^1}|_p, \dots, \partial_{z^d}|_p, \partial_{\bar{z}^1}|_p, \dots, \partial_{\bar{z}^d}|_p\}.$$

From the point of view of real vector spaces, $\partial_{x^j}|_p$ and $i\partial_{x^j}|_p$ would be considered linearly independent and hence $T_p M^{\mathbb{C}}$ has real dimension $4D$.

In exact analogy with the real case, we can define the dual to $T_p M^{\mathbb{C}}$, which we denote by $T_p^* M^{\mathbb{C}} = T_p^* M \otimes \mathbb{C}$, with the one-forms basis

$$T_p^* M^{\mathbb{C}} : \{dz^1|_p, \dots, dz^d|_p, d\bar{z}^1|_p, \dots, d\bar{z}^d|_p\}.$$

For certain types of complex manifolds M , it is worthwhile to refine the definition of the complexified tangent and cotangent spaces, which pulls apart the holomorphic and anti-holomorphic directions in each of these two vector spaces. That is, we can write

$$T_p M^{\mathbb{C}} = T_p M^{(1,0)} \oplus T_p M^{(0,1)},$$

where $T_p M^{(1,0)}$ is the vector space spanned by $\{\partial_{z^1}|_p, \dots, \partial_{z^d}|_p\}$ and $T_p M^{(0,1)}$ is the vector space spanned by $\{\partial_{\bar{z}^1}|_p, \dots, \partial_{\bar{z}^d}|_p\}$. Similarly, we can write

$$T_p^* M^{\mathbb{C}} = T_p^* M^{(1,0)} \oplus T_p^* M^{(0,1)},$$

where $T_p M^{(1,0)}$ is the vector space spanned by $\{dz^1|_p, \dots, dz^d|_p\}$ and $T_p M^{(0,1)}$ is the vector space spanned by $\{d\bar{z}^1|_p, \dots, d\bar{z}^d|_p\}$. We call $T_p M^{(1,0)}$ the *holomorphic tangent space*; it has complex dimension d and we call $T_p^* M^{(1,0)}$ the *holomorphic cotangent space*. It also has complex dimension d . Their complements are known as the *anti-holomorphic* tangent and cotangent spaces respectively [Greene (1996)].

Now, a *complex vector bundle* is a vector bundle $\pi : E \rightarrow M$ whose fiber bundle $\pi^{-1}(x)$ is a complex vector space. It is not necessarily a complex manifold, even if its base manifold M is a complex manifold. If a complex vector bundle also has the structure of a complex manifold, and is holomorphic, then it is called a *holomorphic vector bundle*.

3.14.1 Complex Metrics: Hermitian and Kähler

If M is a complex manifold, there is a natural extension of the metric g to a map

$$g : T_p M^{\mathbb{C}} \times T_p M^{\mathbb{C}} \rightarrow \mathbb{C},$$

defined in the following way. Let r, s, u, v be four vectors in the tangent space $T_p M$ to a complex manifold M . Using them, we can construct, for example, two vectors $w_{(1)} = r + is$ and $w_{(2)} = u + iv$ which lie in $T_p M^{\mathbb{C}}$. Then, we evaluate g on $w_{(1)}$ and $w_{(2)}$ by linearity:

$$g(w_{(1)}, w_{(2)}) = g(r + is, u + iv) = g(r, u) - g(s, v) + i[g(r, v) + g(s, u)] .$$

We can define components of this extension of the original metric (which we have called by the same symbol) with respect to complex coordinates in the usual way: $g_{ij} = g(\frac{\partial}{\partial z^i}, \frac{\partial}{\partial z^j})$, $g_{i\bar{j}} = g(\frac{\partial}{\partial z^i}, \frac{\partial}{\partial \bar{z}^j})$ and so forth. The reality of our original metric g and its symmetry implies that in complex coordinates we have $g_{ij} = g_{ji}$, $g_{i\bar{j}} = g_{\bar{j}i}$ and $\bar{g}_{i\bar{j}} = g_{i\bar{j}}$, $\bar{g}_{i\bar{j}} = g_{i\bar{j}}$.

Now, recall that a *Hermitian metric* on a complex vector bundle assigns a *Hermitian inner product* to every fiber bundle. The basic example is the trivial bundle $\pi : U \times \mathbb{C}^2 \rightarrow U$, where U is an open set in \mathbb{R}^n . Then a positive definite Hermitian matrix H defines a Hermitian metric by

$$\langle v, w \rangle = v^T H \bar{w},$$

where \bar{w} is the complex conjugate of w . By a partition of unity, any complex vector bundle has a Hermitian metric.

In local coordinates of a complex manifold M , a metric g is Hermitian if $g_{ij} = g_{\bar{i}\bar{j}} = 0$. In this case, only the mixed type components of g are nonzero and hence it can be written as

$$g = g_{i\bar{j}} dz^i \otimes d\bar{z}^{\bar{j}} + g_{\bar{i}j} d\bar{z}^{\bar{i}} \otimes dz^j .$$

With a little bit of algebra one can work out the constraint this implies for the original metric written in real coordinates. Formally, if J is a *complex structure* acting on the real tangent space $T_p M$, i.e.

$$J : T_p M \rightarrow T_p M \quad \text{with} \quad J^2 = -I,$$

then the *Hermiticity condition* on g is $g(Jv_{(1)}, Jv_{(2)}) = g(v_{(1)}, v_{(2)})$.

On a holomorphic vector bundle with a Hermitian metric h , there is a unique connection compatible with h and the complex structure. Namely, it must be $\nabla = \partial + \bar{\partial}$.

In the special case of a complex manifold, the complexified tangent bundle $TM \otimes \mathbb{C}$ may have a Hermitian metric, in which case its real part is a Riemannian metric and its imaginary part is a nondegenerate alternating multilinear form ω . When ω is closed, i.e., in this case a symplectic form, then ω is called the *Kähler form*. Formally, given a Hermitian metric g on

M , we can build a form in $\Omega^{1,1}(M)$ — that is, a form of type $(1, 1)$ in the following way:

$$\omega = ig_{i\bar{j}} dz^i \otimes d\bar{z}^{\bar{j}} - ig_{\bar{j}i} d\bar{z}^{\bar{j}} \otimes dz^i.$$

By the symmetry of g , we can write this as

$$\omega = ig_{i\bar{j}} dz^i \wedge d\bar{z}^{\bar{j}}.$$

Now, if ω is closed, that is, if $dJ = 0$, then ω is called a *Kähler form* and M is called a *Kähler manifold*. At first sight, this *Kählerity condition* might not seem too restrictive. However, it leads to remarkable simplifications in the resulting differential geometry on M .

A *Kähler structure* on a complex manifold M combines a Riemannian metric on the underlying real manifold with the complex structure. Such a structure brings together geometry and complex analysis, and the main examples come from algebraic geometry. When M has n complex dimensions, then it has $2n$ real dimensions. A Kähler structure is related to the unitary group $U(n)$, which embeds in $SO(2n)$ as the orthogonal matrices that preserve the *almost complex structure* (multiplication by i). In a coordinate chart, the complex structure of M defines a multiplication by i and the metric defines orthogonality for tangent vectors. On a *Kähler manifold*, these two notions (and their derivatives) are related.

A Kähler manifold is a complex manifold for which the exterior derivative of the fundamental form ω associated with the given *Hermitian metric* vanishes, so $d\omega = 0$. In other words, it is a complex manifold with a *Kähler structure*. It has a *Kähler form*, so it is also a symplectic manifold. It has a *Kähler metric*, so it is also a Riemannian manifold.

The simplest example of a Kähler manifold is a *Riemann surface*, which is a complex manifold of dimension 1. In this case, the imaginary part of any Hermitian metric must be a closed form since all 2-forms are closed on a real 2D manifold.

In other words, a *Kähler form* is a closed two-form ω on a complex manifold M which is also the negative imaginary part of a Hermitian metric $h = g - iw$. In this case, M is called a Kähler manifold and g , the real part of the Hermitian metric, is called a Kähler metric. The Kähler form combines the metric and the complex structure, $g(M, Y) = \omega(M, JY)$, where ω is the *almost complex structure* induced by multiplication by i . Since the Kähler form comes from a Hermitian metric, it is preserved by ω , since $h(M, Y) = h(JX, JY)$. The equation $d\omega = 0$ implies that the metric and

the complex structure are related. It gives M a Kähler structure, and has many implications.

In particular, on \mathbb{C}^2 , the Kähler form can be written as

$$\omega = -\frac{i}{2} (dz_1 \wedge \overline{dz_1} + dz_2 \wedge \overline{dz_2}) = dx_1 \wedge dy_1 + dx_2 \wedge dy_2,$$

where $z_n = x_n + i y_n$. In general, the Kähler form can be written in coordinates

$$\omega = g_{ij} dz_i \wedge \overline{dz_j},$$

where g_{ij} is a Hermitian metric, the real part of which is the Kähler metric. Locally, a Kähler form can be written as $i\partial\bar{\partial}f$, where f is a function called a *Kähler potential*. The Kähler form is a real $(1,1)$ -complex form. The Kähler potential is a real-valued function f on a Kähler manifold for which the Kähler form ω can be written as $\omega = i\partial\bar{\partial}f$, where,

$$\partial = \partial_{z_k} dz_k \quad \text{and} \quad \bar{\partial} = \partial_{\bar{z}_k} d\bar{z}_k.$$

In local coordinates, the fact that $dJ = 0$ for a Kähler manifold M implies

$$dJ = (\partial + \bar{\partial})ig_{i\bar{j}} dz^i \wedge d\bar{z}^{\bar{j}} = 0.$$

This implies that

$$\partial_{z^i} g_{i\bar{j}} = \partial_{z^i} g_{i\bar{j}} \quad (3.240)$$

and similiary with z and \bar{z} interchanged. From this we see that locally we can express $g_{i\bar{j}}$ as

$$g_{i\bar{j}} = \frac{\partial^2 \phi}{\partial z^i \partial \bar{z}^{\bar{j}}}.$$

That is, $\omega = i\partial\bar{\partial}\phi$, where ϕ is a locally defined function in the patch whose local coordinates we are using, which is known as the *Kähler potential*.

If ω on M is a Kähler form, the conditions (3.240) imply that there are numerous cancellations in (3.240). so that the only nonzero *Christoffel symbols* (of the standard *Levi-Civita connection*) in *complex coordinates* are those of the form Γ_{jk}^l and $\Gamma_{\bar{j}\bar{k}}^{\bar{l}}$, with all indices holomorphic or anti-holomorphic. Specifically,

$$\Gamma_{jk}^l = g^{l\bar{s}} \partial_{z^j} g_{k\bar{s}} \quad \text{and} \quad \Gamma_{\bar{j}\bar{k}}^{\bar{l}} = g^{\bar{l}s} \partial_{\bar{z}^{\bar{j}}} g_{\bar{k}s}.$$

The curvature tensor also greatly simplifies. The only non-zero components of the *Riemann tensor*, when written in complex coordinates, have the form $R_{i\bar{j}k\bar{l}}$ (up to index permutations consistent with symmetries of the curvature tensor). And we have

$$R_{i\bar{j}k\bar{l}} = g_{i\bar{s}} \partial_{z^k} \Gamma_{\bar{j}\bar{l}}^{\bar{s}},$$

as well as the *Ricci tensor*

$$R_{\bar{i}j} = R_{\bar{i}\bar{k}j}^{\bar{k}} = -\partial_{z^j} \Gamma_{\bar{i}\bar{k}}^{\bar{k}}.$$

Since the Kähler form ω is closed, it represents a cohomology class in the *de Rham cohomology*. On a compact manifold, it cannot be exact because $\omega^n/n! \neq 0$ is the volume form determined by the metric. In the special case of a projective variety, the Kähler form represents an integral cohomology class. That is, it integrates to an integer on any 1D submanifold, i.e., an algebraic curve. The *Kodaira Embedding Theorem* says that if the Kähler form represents an integral cohomology class on a compact manifold, then it must be a projective variety. There exist Kähler forms which are not projective algebraic, but it is an open question whether or not any Kähler manifold can be deformed to a projective variety (in the compact case).

A Kähler form satisfies *Wirtinger's inequality*,

$$|\omega(M, Y)| \leq |M \wedge Y|,$$

where the r.h.s is the volume of the parallelogram formed by the tangent vectors M and Y . Corresponding inequalities hold for the *exterior powers* of ω . Equality holds iff M and Y form a complex subspace. Therefore, there is a *calibration form*, and the complex submanifolds of a Kähler manifold are *calibrated submanifolds*. In particular, the complex submanifolds are locally volume minimizing in a Kähler manifold. For example, the graph of a holomorphic function is a locally area-minimizing surface in $\mathbb{C}^2 = \mathbb{R}^4$.

Kähler identities is a collection of identities which hold on a Kähler manifold, also called the *Hodge identities*. Let ω be a Kähler form, $d = \partial + \bar{\partial}$ be the exterior derivative, $[A, B] = AB - BA$ be the commutator of two differential operators, and A^* denote the formal adjoint of A . The following operators also act on differential forms α on a Kähler manifold:

$$L(\alpha) = \alpha \wedge \omega, \quad \Lambda(\alpha) = L^*(\alpha) = \alpha \lrcorner \omega, \quad d_c = -JdJ,$$

where J is the almost complex structure, $J = -I$, and \lrcorner denotes the interior

product. Then we have

$$\begin{aligned} [L, \bar{\partial}] &= [L, \partial] = 0, & [\Lambda, \bar{\partial}^*] &= [\Lambda, \partial^*] = 0, \\ [L, \bar{\partial}^*] &= -i\partial, & [L, \partial^*] &= i\bar{\partial}, & [\Lambda, \bar{\partial}] &= -i\partial^*, & [\Lambda, \partial] &= -i\bar{\partial}. \end{aligned}$$

These identities have many implications. For example, the two operators

$$\Delta_d = dd^* + d^*d \quad \text{and} \quad \Delta_{\bar{\partial}} = \bar{\partial}\bar{\partial}^* + \bar{\partial}^*\bar{\partial}$$

(called Laplacians because they are elliptic Laplacian-like operators) satisfy

$$\Delta_d = 2\Delta_{\bar{\partial}}.$$

At this point, assume that M is also a compact manifold. Along with Hodge's Theorem, this equality of Laplacians proves the *Hodge decomposition*. The operators L and Λ commute with these Laplacians. By *Hodge's Theorem*, they act on cohomology, which is represented by *harmonic forms*. Moreover, defining

$$H = [L, \Lambda] = \sum (p + q - n) \Pi^{p,q},$$

where $\Pi^{p,q}$ is projection onto the (p, q) -Dolbeault cohomology, they satisfy

$$[L, \Lambda] = H, \quad [H, L] = -2L, \quad [H, \Lambda] = 2L.$$

In other words, these operators give a group representation of the special linear Lie algebra $\mathfrak{sl}_2(\mathbb{C})$ on the complex cohomology of a compact Kähler manifold (*Lefschetz Theorem*).

3.14.2 Calabi–Yau Manifolds

A *Calabi–Yau manifold* is a Kähler manifold of complex dimension n with a covariant constant holomorphic n -form. Equivalently it is a Riemannian manifold with holonomy contained in $SU(n)$.

It is convenient for our purposes to play down the role of the complex structure in describing such manifolds and to emphasize instead the role of three closed forms, satisfying certain algebraic identities. We have the Kähler 2-form ω and the real and imaginary parts Ω_1 and Ω_2 of the covariant constant n -form. These satisfy some identities [Hitchin (1997); Atiyah and Hitchin (1988)]:

- (i) ω is non-degenerate
- (ii) $\Omega_1 + i\Omega_2$ is locally decomposable and non-vanishing
- (iii) $\Omega_1 \wedge \omega = \Omega_2 \wedge \omega = 0$

- (iv) $(\Omega_1 + i\Omega_2) \wedge (\Omega_1 - i\Omega_2) = \omega^n$ (resp. $i\omega^n$) if n is even (resp. odd)
 (v) $d\omega = 0$, $d\Omega_1 = 0$, $d\Omega_2 = 0$.

These conditions (together with a positivity condition) we now show serve to characterize CY manifolds. Firstly if $\Omega^c = \Omega_1 + i\Omega_2$ is locally decomposable as $\theta_1 \wedge \theta_2 \wedge \dots \wedge \theta_n$, then take the subbundle Λ of $T^*M \otimes \mathbb{C}$ spanned by $\theta_1, \dots, \theta_n$. By (iv) and the fact that $\omega^n \neq 0$, we have

$$\theta_1 \wedge \dots \wedge \theta_n \wedge \bar{\theta}_1 \wedge \dots \wedge \bar{\theta}_n \neq 0$$

and so $T^*M = \Lambda + \bar{\Lambda}$ and we have an almost-complex structure. In this description a 1-form θ is of type $(1, 0)$ iff $\Omega^c \wedge \theta = 0$. Since from (v) $d\Omega_1 = d\Omega_2 = 0$, this means that $\Omega^c \wedge d\theta = 0$. Writing

$$d\theta = \sum a_{ij}\theta_i \wedge \theta_j + \sum b_{ij}\theta_i \wedge \bar{\theta}_j + \sum c_{ij}\bar{\theta}_i \wedge \bar{\theta}_j \quad (3.241)$$

we see that $c_{ij} = 0$. Thus the ideal generated by Λ is closed under exterior differentiation, and (by the Newlander–Nirenberg Theorem) the structure is integrable.

Similarly, applying the decomposition of 2-forms (3.241) to ω , (iii) implies that the $(0, 2)$ -component vanishes, and since ω is real, it is of type $(1, 1)$. It is closed by (v), so if the Hermitian form so defined is positive definite, then we have a Kähler metric.

Since Ω^c is closed and of type $(n, 0)$ it is a non-vanishing holomorphic section s of the canonical bundle. Relative to the trivialization s , the Hermitian connection has connection form given by $\partial \ln(\|s\|^2)$. But property (iv) implies that it has constant length, so the connection form vanishes and $s = \Omega^c$ is covariant constant.

3.14.3 Special Lagrangian Submanifolds

A submanifold L of a symplectic manifold X is Lagrangian if ω restricts to zero on L and $\dim X = 2 \dim L$. A submanifold of a CY manifold is *special Lagrangian* if in addition $\Omega = \Omega_1$ restricts to zero on L . This condition involves only two out of the three forms, and in many respects what we shall be doing is to treat them both, the 2-form ω and the n -form Ω , on the same footing.

Here we need to emphasize the following remarks:

1. We could relax the definition a little since Ω^c is a *chosen* holomorphic n -form: any constant multiple of Ω^c would also be covariant constant, so under some circumstances we may need to say that L is special Lagrangian if, for some non-zero $c_1, c_2 \in \mathbb{R}$, $c_1\Omega_1 + c_2\Omega_2 = 0$.

2. On a special Lagrangian submanifold L , the n -form Ω_2 restricts to a non-vanishing form, so in particular L is always oriented.

Examples of special Lagrangian submanifolds are difficult to find, and so far consist of three types [Hitchin (1997); Atiyah and Hitchin (1988)]:

- Complex Lagrangian submanifolds of hyperkähler manifolds;
- Fixed points of a real structure on a CY manifold;
- Explicit examples for non-compact CY manifolds;

The hyperkähler examples arise easily. In this case we have $n = 2k$ and three Kähler forms $\omega_1, \omega_2, \omega_3$ corresponding to the three complex structures I, J, K of the hyperkähler manifold. With respect to the complex structure I the form $\omega^c = (\omega_2 + i\omega_3)$ is a holomorphic symplectic form. If L is a complex Lagrangian submanifold (i.e. L is a complex submanifold and ω^c vanishes on L), then the real and imaginary parts of this, ω_2 and ω_3 , vanish on L . Thus $\omega = \omega_2$ vanishes and if k is odd (resp. even), the real (resp. imaginary) part of $\Omega^c = (\omega_3 + i\omega_1)^k$ vanishes. Using the complex structure J instead of I , we see that L is special Lagrangian. For examples here, we can take any holomorphic curve in a K3 surface S , or its symmetric product in the Hilbert scheme $S^{[m]}$, which is hyperkähler from [Hitchin (1997); Atiyah and Hitchin (1988)].

If X is a CY manifold with a real structure — an antiholomorphic involution σ , for which $\sigma^*\omega = -\omega$ and $\sigma^*\Omega = -\Omega$, then the fixed point set (the set of real points of X) is easily seen to be a special Lagrangian submanifold L . All CY metrics on compact manifolds are produced by the existence Theorem of Yau. In particular T^*S^n (with the complex structure of an affine quadric) has a complete CY metric for which the zero section is special Lagrangian.

3.14.4 Dolbeault Cohomology and Hodge Numbers

A generalization of the real-valued de Rham cohomology to complex manifolds is called the *Dolbeault cohomology*. On complex m D manifolds, we have local coordinates z^i and \bar{z}^i . One can now study (p, q) -forms, which are forms containing p factors of dz^i and q factors of $d\bar{z}^j$:

$$\omega = \omega_{i_1 \dots i_p, j_1 \dots j_q}(z, \bar{z}) dz^{i_1} \wedge \dots \wedge dz^{i_p} \wedge d\bar{z}^{j_1} \wedge \dots \wedge d\bar{z}^{j_q}.$$

Moreover, one can introduce *two* exterior derivative operators ∂ and $\bar{\partial}$,

where ∂ is defined by

$$\partial\omega \equiv \frac{\partial\omega_{i_1\dots i_p, j_1\dots j_q}}{\partial z^k} dz^k \wedge dz^{i_1} \wedge \dots \wedge dz^{i_p} \wedge d\bar{z}^{j_1} \wedge \dots \wedge d\bar{z}^{j_q},$$

and $\bar{\partial}$ is defined similarly by differentiating with respect to \bar{z}^k and adding a factor of $d\bar{z}^k$. Again, both of these operators square to zero. We can now construct two cohomologies – one for each of these operators – but as we will see, in the cases that we are interested in, the information contained in them is the same. Conventionally, one uses the cohomology defined by the $\bar{\partial}$ –operator.

For complex manifolds, the *Hodge Theorem* also holds: *each cohomology class $H^{p,q}(M)$ contains a unique harmonic form*. Here, a harmonic form ω_h is a form for which the *complex Laplacian*

$$\Delta_{\bar{\partial}} = \bar{\partial}\bar{\partial}^* + \bar{\partial}^*\bar{\partial}$$

has a zero eigenvalue: $\Delta_{\bar{\partial}}\omega_h = 0$. In general, this operator does not equal the ordinary Laplacian, but one can prove that in the case where M is a *Kähler manifold*,

$$\Delta = 2\Delta_{\bar{\partial}} = 2\Delta_{\partial}.$$

In other words, on a Kähler manifold the notion of a harmonic form is the same, independently of which exterior derivative one uses. As a first consequence, we find that the vector spaces $H_{\bar{\partial}}^{p,q}(M)$ and $H_{\partial}^{p,q}(M)$ both equal the vector space of harmonic (p, q) –forms, so the two cohomologies are indeed equal. Moreover, every (p, q) –form is a $(p + q)$ –form in the de Rham cohomology, and by the above result we see that a *harmonic* (p, q) –form can also be viewed as a de Rham harmonic $(p + q)$ –form. Conversely, any de Rham p –form can be written as a sum of Dolbeault forms:

$$\omega_p = \omega_{p,0} + \omega_{p-1,1} + \dots + \omega_{0,p}. \quad (3.242)$$

Acting on this with the Laplacian, one sees that for a harmonic p –form,

$$\Delta\omega_p = \Delta_{\bar{\partial}}\omega_p = \Delta_{\bar{\partial}}\omega_{p,0} + \Delta_{\bar{\partial}}\omega_{p-1,1} + \dots + \Delta_{\bar{\partial}}\omega_{0,p} = 0.$$

Since $\Delta_{\bar{\partial}}$ does not change the degree of a form, $\Delta_{\bar{\partial}}\omega_{p_1,p_2}$ is also a (p_1, p_2) –form. Therefore, the r.h.s. can only vanish if each term vanishes separately, so all the terms on the r.h.s. of (3.242) must be harmonic forms. Summarizing, we have shown that the vector space of harmonic de Rham p –forms is a direct sum of the vector spaces of harmonic Dolbeault

(p_1, p_2) –forms with $p_1 + p_2 = p$. Since the harmonic forms represent the cohomology classes in a 1–1 way, we find the important result that for Kähler manifolds,

$$H^p(M) = H^{p,0}(M) \oplus H^{p-1,1}(M) \oplus \dots \oplus H^{0,p}(M).$$

That is, the Dolbeault cohomology can be viewed as a refinement of the de Rham cohomology. In particular, we have

$$b^p = h^{p,0} + h^{p-1,1} + \dots + h^{0,p},$$

where $h^{p,q} = \dim H^{p,q}(M)$ are called the *Hodge numbers* of M .

The Hodge numbers of a Kähler manifold give us several topological invariants, but not all of them are independent. In particular, the following two relations hold:

$$h^{p,q} = h^{q,p}, \quad h^{p,q} = h^{m-p, m-q}. \quad (3.243)$$

The first relation immediately follows if we realize that $\omega \mapsto \bar{\omega}$ maps ∂ –harmonic (p, q) –forms to $\bar{\partial}$ –harmonic (q, p) –forms, and hence can be viewed as an invertible map between the two respective cohomologies. As we have seen, the ∂ –cohomology and the $\bar{\partial}$ –cohomology coincide on a Kähler manifold, so the first of the above two equations follows.

The second relation can be proved using the map

$$(\alpha, \omega) \mapsto \int_M \alpha \wedge \omega$$

from $H^{p,q} \times H^{m-p, m-q}$ to \mathbb{C} . It can be shown that this map is nondegenerate, and hence that $H^{p,q}$ and $H^{m-p, m-q}$ can be viewed as dual vector spaces. In particular, it follows that these vector spaces have the same dimension, which is the statement in the second line of (3.243).

Note that the last argument also holds for de Rham cohomology, in which case we find the relation $b^p = b^{n-p}$ between the Betti numbers. We also know that $H^{n-p}(M)$ is dual to $H_{n-p}(M)$, so combining these statements we find an identification between the vector spaces $H^p(M)$ and $H_{n-p}(M)$. Recall that this identification between p –form cohomology classes and $(n-p)$ –cycle homology classes represents the *Poincaré duality*. Intuitively, take a certain $(n-p)$ –cycle Σ representing a homology class in H_{n-p} . One can now try to define a ‘delta function’ $\delta(\Sigma)$ which is localized on this cycle. Locally, Σ can be parameterized by setting p coordinates equal to zero, so $\delta(\Sigma)$ is a ‘ p D delta function’ – that is, it is an object which is naturally integrated over p D submanifolds: a p –form. This

intuition can be made precise, and one can indeed view the cohomology class of the resulting ‘delta-function’ p -form as the Poincaré dual to Σ .

Going back to the relations (3.243), we see that the Hodge numbers of a Kähler manifold can be nicely written in a so-called *Hodge diamond* form:

$$\begin{array}{ccccc}
 & & h^{0,0} & & \\
 & h^{1,0} & & h^{0,1} & \\
 & \ddots & & \ddots & \\
 h^{m,0} & & \dots & & h^{0,m} \\
 & \ddots & & \ddots & \\
 & h^{m,m-1} & & h^{m-1,m} & \\
 & & h^{m,m} & &
 \end{array}$$

The integers in this diamond are symmetrical under the reflection in its horizontal and vertical axes.

3.15 Conformal Killing–Riemannian Geometry

In this section we present some basic facts from *conformal Killing–Riemannian geometry*. In mechanics it is well-known that *symmetries of Lagrangian* or *Hamiltonian* result in *conservation laws*, that are used to deduce constants of motion for the trajectories (geodesics) on the configuration manifold M . The same constants of motion are get using geometrical language, where a *Killing vector-field* is the standard tool for the description of symmetry [Misner *et al.* (1973)]. A Killing vector-field ξ^i is a vector-field on a Riemannian manifold M with metrics g , which in coordinates $x^j \in M$ satisfies the *Killing equation*

$$\xi^{i;j} + \xi^{j;i} = \xi^{(i;j)} = 0, \quad \text{or} \quad \mathcal{L}_{\xi^i} g_{ij} = 0, \quad (3.244)$$

where semicolon denotes the covariant derivative on M , the indexed bracket denotes the tensor symmetry, and \mathcal{L} is the Lie derivative.

The *conformal Killing vector-fields* are, by definition, infinitesimal conformal symmetries i.e., the flow of such vector-fields preserves the conformal class of the metric. The number of linearly-independent conformal Killing fields measures the degree of *conformal symmetry of the manifold*. This number is bounded by $\frac{1}{2}(n+1)(n+2)$, where n is the dimension of the manifold. It is the maximal one if the manifold is conformally flat [Baum (2000)].

Now, to properly initialize our conformal geometry, recall that *conformal twistor spinor-fields* φ were introduced by R. Penrose into physics (see [Penrose (1967); Penrose and Rindler (1984); Penrose and Rindler (1986)]) as solutions of the *conformally covariant twistor equation*

$$\nabla_X^S \varphi + \frac{1}{n} X \cdot D\varphi = 0,$$

for each vector-fields X on a Riemannian manifold (M, g) , where D is the *Dirac operator*. Each twistor spinor-field φ on (M, g) defines a *conformal vector-field* V_φ on M by

$$g(V_\varphi, X) = i^{k+1} \langle X \cdot \varphi, \varphi \rangle.$$

Also, each twistor spinor-field φ that satisfies the *Dirac equation* on (M, g) ,

$$D\varphi = \mu\varphi,$$

is called a *Killing spinor-field*. Each twistor spinor-field without zeros on (M, g) can be transformed by a conformal change of the metric g into a Killing spinor-field [Baum (2000)].

3.15.1 Conformal Killing Vector-Fields and Forms on M

The space of all conformal Killing vector-fields forms the Lie algebra of the *isometry group* of a Riemannian manifold (M, g) and the number of linearly independent Killing vector-fields measures the *degree of symmetry* of M . It is known that this number is bounded from above by the dimension of the isometry group of the standard sphere and, on compact manifolds, equality is attained if and only if the manifold M is isometric to the standard sphere or the real projective space. Slightly more generally one can consider *conformal vector-fields*, i.e., vector-fields with a flow preserving a given conformal class of metrics. There are several geometrical conditions which force a conformal vector-field to be Killing [Simmelmenn (2002)].

A natural generalization of conformal vector-fields are the *conformal Killing forms* [Yano (1952)], also called *twistor forms* [Moroianu and Semmelmann (2003)]. These are p -forms α satisfying for any vector-field X on the manifold M the *Killing-Yano equation*

$$\nabla_X \alpha - \frac{1}{p+1} X \lrcorner d\alpha + \frac{1}{n-p+1} X^* \wedge d^* \alpha = 0, \quad (3.245)$$

where n is the dimension of the manifold (M, g) , ∇ denotes the covariant derivative of the Levi-Civita connection on M , X^* is 1-form dual to X and

\rfloor is the operation dual to the wedge product on M . It is easy to see that a conformal Killing 1-form is dual to a conformal vector-field. *Coclosed* conformal Killing p -forms are called *Killing forms*. For $p = 1$ they are dual to Killing vector-fields.

Let α be a Killing p -form and let γ be a geodesic on (M, g) , i.e., $\nabla_{\dot{\gamma}} \dot{\gamma} = 0$. Then

$$\nabla_{\dot{\gamma}} (\dot{\gamma} \rfloor \alpha) = (\nabla_{\dot{\gamma}} \dot{\gamma}) \rfloor \alpha + \dot{\gamma} \rfloor \nabla_{\dot{\gamma}} \alpha = 0,$$

i.e., $\dot{\gamma} \rfloor \alpha$ is a $(p-1)$ -form parallel along the geodesic γ and in particular its length is constant along γ .

The l.h.s of equation (3.245) defines a first-order elliptic differential operator T , the so-called *twistor operator*. Equivalently one can describe a conformal Killing form as a form in the kernel of twistor operator T . From this point of view conformal Killing forms are similar to Penrose's *twistor spinors* in *Lorentzian spin geometry*. One shared property is the conformal invariance of the defining equation. In particular, any form which is parallel for some metric g , and thus a Killing form for trivial reasons, induces non-parallel conformal Killing forms for metrics conformally equivalent to g (by a non-trivial change of the metric) [Simmelmann (2002)].

3.15.2 Conformal Killing Tensors and Laplacian Symmetry

In an n D Riemannian manifold (M, g) , a *Killing tensor-field* (of order 2) is a symmetric tensor K^{ab} satisfying (generalizing (3.244))

$$K^{(ab;c)} = 0. \quad (3.246)$$

A *conformal Killing tensor-field* (of order 2) is a symmetric tensor Q^{ab} satisfying

$$Q^{(ab;c)} = q^{(a} g^{bc)}, \quad \text{with} \quad q^a = (Q^{,a} + 2Q_d^{a;d})/(n+2), \quad (3.247)$$

where comma denotes partial derivative and $Q = Q_d^d$. When the associated *conformal vector* q^a is nonzero, the conformal Killing tensor will be called *proper* and otherwise it is a (ordinary) Killing tensor. If q^a is a Killing vector, Q^{ab} is referred to as a *homothetic Killing tensor*. If the associated conformal vector $q^a = q^{,a}$ is the gradient of some scalar field q , then Q^{ab} is called a *gradient conformal Killing tensor*. For each gradient conformal

Killing tensor Q^{ab} there is an associated Killing tensor K^{ab} given by

$$K^{ab} = Q^{ab} - qg^{ab}, \quad (3.248)$$

which is defined only up to the addition of a constant multiple of the inverse metric tensor g^{ab} .

Some authors define a conformal Killing tensor as a *trace-free* tensor P^{ab} satisfying $P^{(ab;c)} = p^{(a}g^{bc)}$. Note that there is no contradiction between the two definitions: if P^{ab} is a trace-free conformal Killing tensor then for any scalar field λ , $P^{ab} + \lambda g^{ab}$ is a conformal Killing tensor and conversely if Q^{ab} is a conformal Killing tensor, its trace-free part $Q^{ab} - \frac{1}{n}Qg^{ab}$ is a trace-free Killing tensor [Rani *et. al.* (2003)].

Killing tensor-fields are of importance owing to their connection with quadratic first integrals of the geodesic equations: if p^a is tangent to an affinely parameterized geodesic (i.e., $p^a{}_{;b}p^b = 0$) it is easy to see that $K_{ab}p^ap^b$ is constant along the geodesic. For conformal Killing tensors $Q_{ab}p^ap^b$ is constant along *null* geodesics and here, only the trace-free part of Q_{ab} contributes to the constants of motion. Both Killing tensors and conformal Killing tensors are also of importance in connection with the separability of the Hamiltonian-Jacobi equations [Conway and Hopf (1964)] (as well as other PDEs).

A Killing tensor is said to be *reducible* if it can be written as a constant linear combination of the metric and symmetrized products of Killing vectors,

$$K_{ab} = a^0g_{ab} + a^{IJ}\xi_{I(a}\xi_{|J|b)}, \quad (3.249)$$

where ξ_I for $I = 1 \dots N$ are the Killing vectors admitted by the manifold (M, g) and a^0 and a^{IJ} for $1 \leq I \leq J \leq N$ are constants. Generally one is interested only in Killing tensors which are not reducible since the quadratic constant of motion associated with a reducible Killing tensor is a constant linear combination of p^ap_a and of pairwise products of the linear constants of motion $\xi_{Ia}p^a$ [Rani *et. al.* (2003)].

More generally, any linear differential operator on a Riemannian manifold (M, g) may be written in the form [Eastwood (1991); Eastwood (2002)]

$$\mathcal{D} = V^{bc\dots d}\nabla_b\nabla_c\dots\nabla_d + \text{lower order terms},$$

where $V^{bc\dots d}$ is symmetric in its indices, and $\nabla_a = \partial/\partial x^a$ (differentiation in coordinates). This tensor is called the symbol of \mathcal{D} . We shall write $\phi^{(ab\dots c)}$ for the symmetric part of $\phi^{ab\dots c}$.

Now, a *conformal Killing tensor* on (M, g) is a symmetric trace-free tensor-field, with s indices, satisfying

$$\text{the trace-free part of } \nabla^{(a} V^{bc\dots d)} = 0, \quad (3.250)$$

or, equivalently,

$$\nabla^{(a} V^{bc\dots d)} = g^{(ab} T^{c\dots d)}, \quad (3.251)$$

for some tensor-field $T^{c\dots d}$ or, equivalently,

$$\nabla^{(a} V^{bc\dots d)} = \frac{s}{n+2s-2} g^{(ab} \nabla_e V^{c\dots d)e}, \quad (3.252)$$

where $\nabla^a = g^{ab} \nabla_b$ (the standard convention of raising and lowering indices with the metric tensor g_{ab}). When $s = 1$, these equations define a *conformal Killing vector*.

M. Eastwood proved the following Theorem: any symmetry \mathcal{D} of the Laplacian $\Delta = \nabla^a \nabla_a$ on a Riemannian manifold (M, g) is canonically equivalent to one whose symbol is a conformal Killing tensor [Eastwood (1991); Eastwood (2002)].

3.15.3 Application: Killing Vector and Tensor Fields in Mechanics

Recall from subsection 3.15 above, that on a Riemannian manifold (M, g) with the system's kinetic energy metric tensor $g = (g_{ij})$, for any pair of vectors V and T , the following relation holds⁹

$$\partial_s \langle V, T \rangle = \langle \nabla_s V, T \rangle + \langle V, \nabla_s T \rangle, \quad (3.253)$$

where $\langle V, T \rangle = g_{ij} V^i T^j$. If the curve $\gamma(s)$ is a geodesic, for a generic vector X we have

$$\partial_s \langle X, \dot{\gamma} \rangle = \langle \nabla_s X, \dot{\gamma} \rangle + \langle X, \nabla_s \dot{\gamma} \rangle = \langle \nabla_s X, \dot{\gamma} \rangle \equiv \langle \nabla_{\dot{\gamma}} X, \dot{\gamma} \rangle, \quad (3.254)$$

where

$$(\nabla_{\dot{\gamma}} X)^i = \partial_s x^l \partial_{x^l} X^i + \Gamma_{jk}^i \partial_s x^j X^k,$$

so that in components it reads

$$\partial_s (X_i \dot{x}^i) = \dot{x}^i \nabla_i (X_j \dot{x}^j).$$

⁹In this subsection, the overdot denotes the derivative upon the arc-length parameter s , namely $\dot{(\)} \equiv \partial_s \equiv d/ds$, while ∇_s is the covariant derivative along a curve $\gamma(s)$.

Using the fact that $X_j \dot{x}^i \nabla_i \dot{x}^j = X_j \nabla_{\dot{\gamma}} \dot{\gamma}^j = 0$, as well as the auto-parallelism of the geodesics, this can be rewritten as

$$\partial_s(X_i \dot{x}^i) = \frac{1}{2} \dot{x}^j \dot{x}^i (\nabla_i X_j + \nabla_j X_i), \quad (i, j = 1, \dots, N).$$

This means that the conservation of $X_i \dot{x}^i$ along a geodesic, i.e., $\partial_s(X_i \dot{x}^i) = 0$, is guaranteed by (see [Clementi and Pettini (2002)])

$$\nabla_{(i} X_{j)} \equiv \nabla_i X_j + \nabla_j X_i = 0. \quad (3.255)$$

If such a field exists on a manifold, it is the *Killing vector-field*. Recall that (3.255) is equivalent to $\mathcal{L}_X g = 0$, where \mathcal{L} is the *Lie derivative*. On the biodynamical manifolds (M, g) , being the unit vector \dot{q}^i – tangent to a geodesic – proportional to the canonical momentum $p_i = \frac{\partial L}{\partial \dot{q}^i} = \dot{q}^i$, the existence of a Killing vector-field X implies that the corresponding *momentum map* (see subsection 3.12.3.5 above),

$$J(q, p) = X_k(q) \partial_s q^k = \frac{1}{\sqrt{2}(E - V(q))} X_k(q) \dot{q}^k = \frac{1}{\sqrt{2}T(q)} X_k(q) p_k, \quad (3.256)$$

is a constant of motion along the geodesic flow. Thus, for an NDOF Hamiltonian system, a physical conservation law, involving a conserved quantity linear in the canonical momenta, can always be related with a symmetry on the manifold (M, g) due to the action of a Killing vector-field on the manifold. These are the *Noether conservation laws*. The equation (3.255) is equivalent to the vanishing of the Poisson brackets,

$$\{H, J\} = \left(\frac{\partial H}{\partial q^i} \frac{\partial J}{\partial p_i} - \frac{\partial H}{\partial p_i} \frac{\partial J}{\partial q^i} \right) = 0, \quad (3.257)$$

which is the standard definition of a constant of motion $J(q, p)$ (see, e.g., [Abraham and Marsden (1978)]).

However, if a 1–1 correspondence is to exist between conserved physical quantities along a Hamiltonian flow and suitable symmetries of the biodynamical manifolds (M, g) , then *integrability* will be equivalent to the existence of a number of symmetries at least equal to the number of DOF, which is equal to $\dim(M)$. If a Lie group G acts on the phase-space manifold through completely canonical transformations, and there exists an associated momentum map, then every Hamiltonian having G as a symmetry group, with respect to its action, admits the momentum map as the constant of motion. These symmetries are usually referred to as *hidden*

symmetries because, even though their existence is ensured by integrability, they are not easily recognizable [Clementi and Pettini (2002)].

Let us now extend what has been presented so far about Killing vector-fields, trying to generalize the form of the conserved quantity along a geodesic flow from $J = X_i \dot{x}^i$ to $J = K_{j_1 j_2 \dots j_r} \dot{x}^{j_1} \dot{x}^{j_2} \dots \dot{x}^{j_r}$, with $K_{j_1 j_2 \dots j_r}$ a tensor of rank r . Thus, we look for the conditions that entail

$$\partial_s(K_{j_1 j_2 \dots j_r} \dot{x}^{j_1} \dot{x}^{j_2} \dots \dot{x}^{j_r}) = \dot{x}^j \nabla_j (K_{j_1 j_2 \dots j_r} \dot{x}^{j_1} \dot{x}^{j_2} \dots \dot{x}^{j_r}) = 0. \quad (3.258)$$

In order to work out from this equation a condition for the existence of a suitable tensor $K_{j_1 j_2 \dots j_r}$, which is called a *Killing tensor-field*, let us first consider the $2r$ rank tensor $K_{j_1 j_2 \dots j_r} \dot{x}^{i_1} \dot{x}^{i_2} \dots \dot{x}^{i_r}$ and its covariant derivative along a geodesic [Clementi and Pettini (2002)]

$$\dot{x}^j \nabla_j (K_{j_1 j_2 \dots j_r} \dot{x}^{i_1} \dot{x}^{i_2} \dots \dot{x}^{i_r}) = \dot{x}^{i_1} \dot{x}^{i_2} \dots \dot{x}^{i_r} \dot{x}^j \nabla_j K_{j_1 j_2 \dots j_r}, \quad (3.259)$$

where we have again used $\dot{x}^j \nabla_j \dot{x}^{i_k} = 0$ along a geodesic, and a standard covariant differentiation formula (see 3.10.1 above). Now, by contraction on the indices i_k and j_k the $2r$ -rank tensor in (3.259) gives a new expression for (3.258), which reads

$$\partial_s(K_{j_1 j_2 \dots j_r} \dot{x}^{j_1} \dot{x}^{j_2} \dots \dot{x}^{j_r}) = \dot{x}^{j_1} \dot{x}^{j_2} \dots \dot{x}^{j_r} \dot{x}^j \nabla_j K_{j_1 j_2 \dots j_r}, \quad (3.260)$$

where $\nabla_{(j} K_{j_1 j_2 \dots j_r)} = \nabla_j K_{j_1 j_2 \dots j_r} + \nabla_{j_1} K_{j j_2 \dots j_r} + \dots + \nabla_{j_r} K_{j_1 j_2 \dots j_{r-1} j}$. The vanishing of (3.260), entailing the conservation of $K_{j_1 j_2 \dots j_r} \dot{x}^{j_1} \dot{x}^{j_2} \dots \dot{x}^{j_r}$ along a geodesic flow, is therefore guaranteed by the existence of a tensor-field fulfilling the conditions [Clementi and Pettini (2002)]

$$\nabla_{(j} K_{j_1 j_2 \dots j_r)} = 0. \quad (3.261)$$

These equations generalize (3.255) and give the definition of a Killing tensor-field on a Riemannian biodynamical manifold (M, g) . These N^{r+1} equations in $(N + r - 1)!/r!(N - 1)!$ unknown independent components of the Killing tensor constitute an *overdetermined* system of equations. Thus, *a priori*, we can expect that the existence of Killing tensor-fields has to be rather exceptional.

If a Killing tensor-field exists on a Riemannian manifold, then the scalar

$$K_{j_1 j_2 \dots j_r} \dot{q}^{j_1} \dot{q}^{j_2} \dots \dot{q}^{j_r}$$

is a constant of motion for the geodesic flow on the same manifold. With the only difference of a more tedious combinatorics, also in this case it turns

out that the equations (3.261) are equivalent to the vanishing of the Poisson brackets of $J(q, p)$, that is

$$\{H, J\} = 0 \quad \text{is equivalent to} \quad \nabla_{(j} K_{j_1 j_2 \dots j_r)} = 0.$$

Thus, the existence of Killing tensor-fields, obeying (3.261), on a biodynamical manifold (M, g) give the rephrasing of integrability of Newtonian equations of motion or, equivalently, of standard Hamiltonian systems, within the Riemannian geometrical framework.

The first natural question to address concerns the existence of a Killing tensor-field, on any biodynamical manifold (M, g) , to be associated with total energy conservation. Such a Killing tensor-field actually exists and coincides with the metric tensor g , in fact it satisfies by definition (3.261).

One of the simplest case of integrable system is represented by a decoupled system described by a generic Hamiltonian

$$H = \sum_{i=1}^N \left[\frac{p_i^2}{2} + V_i(q^i) \right] = \sum_{i=1}^N H_i(q^i, p_i)$$

for which all the energies E_i of the subsystems H_i , $i = 1, \dots, N$, are conserved. On the associated biodynamical manifold, N second-order Killing tensor-fields exist, they are given by

$$K_{jk}^{(i)} = \delta_{jk} \{V_i(q^i)[E - V(q^i)] + \delta_j^i [E - V(q^i)]^2\}.$$

In fact, these tensor-fields fulfil (3.261), which explicitly reads [Clementi and Pettini (2002)]

$$\begin{aligned} & \nabla_k K_{lm}^{(i)} + \nabla_l K_{mk}^{(i)} + \nabla_m K_{kl}^{(i)} \\ &= \partial_{q^k} K_{lm}^{(i)} + \partial_{q^l} K_{mk}^{(i)} + \partial_{q^m} K_{kl}^{(i)} - 2\Gamma_{kl}^j K_{jm}^{(i)} - 2\Gamma_{km}^j K_{jl}^{(i)} - 2\Gamma_{lm}^j K_{jk}^{(i)} = 0. \end{aligned}$$

The conserved quantities $J^{(i)}(q, p)$ are then get by saturation of the tensors $K^{(i)}$ with the velocities \dot{q}^i ,

$$J^{(i)}(q, p) = K_{jk}^{(i)} \dot{q}^j \dot{q}^k = E_i.$$

3.16 Application: Lax-Pair Tensors in Gravitation

Recall that many problems in *general relativity* require an understanding of the global structure of the space-time. Currently discussed global problems include the occurrence of naked singularities [Ori and Piran (1987)] and universality in gravitational collapse situations [Choptuik (1993)]. The

study of global properties of space-times relies to a large extent on the ability to integrate the geodesic equations. In the absence of exact solutions numerical integration is often used to get a quantitative picture. However, in the quest for a deeper understanding the exact and numerical approaches should be viewed as complementary tools. To perform an exact investigation of the global properties of a given space-time, not only must the space-time itself be an exact solution of the Einstein equations, but in addition the geodesic equations must be integrable. Usually, in a d -dimensional space, integrability of the geodesic equations is connected with the existence of at least $d - 1$ mutually commuting Killing vector fields which span a hypersurface in the space-time. There are exceptions however. The most well-known example is the Kerr space-time which has only two commuting Killing vectors. In that case it is the existence of an irreducible second rank Killing tensor which makes integration possible [Walker and Penrose (1970)]. Another example is given by Ozsvath's class III cosmologies [Ozsvath (1970)]. In that case the geodesic system was integrated using the existence of a non-Abelian Lie algebra of Killing vectors [Rosquist (1980)]. In general integrability can only be guaranteed if there is a set of d constants of the motion in involution (i.e. mutually Poisson commuting). Since the metric itself always provides one constant of the motion corresponding to the squared length of the geodesic tangent vector, the geodesic system will be integrable by Liouville's theorem if there are $d - 1$ additional Poisson commuting invariants.

Exact solutions of Einstein's equations typically admit a number of Killing vector fields. Some of these Killing vector fields may be motivated by physical considerations. For example if one is interested in static stars the space-time must have a time-like Killing vector. For such systems it is also very reasonable to assume spherical spatial symmetry leading to a total of four (non-commuting) Killing vectors. In most cases, the number of Killing vectors is limited by the physics of the problem. In a spherically symmetric collapse situation, for example, the space-time admits exactly three non-commuting Killing vector fields which form an isometry group with 2D orbits. That structure is not sufficient for an exact integration of the geodesic equations. However, the physics of the problem does not impose any *a priori* restrictions on higher rank (≥ 2) Killing tensors. A Killing vector field, ξ , plays a double role; it is both an isometry for the metric ($\mathcal{L}_\xi g = 0$) and a geodesic symmetry. This last property means that it can be interpreted as a symmetry transformation for the geodesic equations. By contrast higher rank Killing tensors are only geodesic symmetries. They

have no obvious geometric interpretation ([Rosquist (1989)]). Because of the isometry property of the Killing vector fields, such symmetries can be incorporated right from the start by assuming a particular form the metric. In this way the field equations are actually simplified by the assumption of Killing vector symmetries. On the other hand, the higher rank Killing symmetries can at present not be used to simplify the form of the field equations. Instead the Killing tensor equations must be imposed as extra conditions thereby increasing both the number of dependent variables and the number of equations.

The *Lax tensors* introduced in [Rosquist (1997)] provide a unifying framework for Killing tensors of any rank and may lead to possibilities to incorporate the higher Killing symmetries in the field equations themselves. A single Lax tensor may generate Killing tensors of varying ranks. Lax tensors arise from a covariant formulation of the *Lax pair equation* [Lax (1968)] for Riemannian and pseudo-Riemannian geometries. The standard Lax pair formulation involves a pair of matrices. In the covariant formulation on the other hand, the Lax pair is represented by two third rank tensors. The first Lax matrix corresponds exactly to the first Lax tensor while the second Lax matrix and the second Lax tensor differ by a term which coincides with the *Levi-Civita connection*. The derivative part of the tensorial Lax pair equation is identical to the *Killing-Yano equation*. Therefore *Killing-Yano tensors* are special cases of Lax tensors for which the second Lax tensor vanishes (the second Lax matrix however does not). However, whereas Killing-Yano tensors are by definition totally antisymmetric the Lax tensors have no *a priori* symmetry restrictions.

3.16.1 *Lax-Pair Tensors*

In this subsection we outline the approach to integrable geometries as given in [Rosquist (1997)]. We consider a Riemannian or pseudo-Riemannian manifold with metric

$$ds^2 = g_{\mu\nu} dq^\mu dq^\nu.$$

The geodesic equations can be represented by the Hamiltonian

$$H = \frac{1}{2} g^{\mu\nu} p_\mu p_\nu, \quad (3.262)$$

together with the natural *Poisson bracket* (denoted by $\{, \}$) on the cotangent bundle. The geodesic system is given by

$$\dot{q}^\alpha = \{q^\alpha, H\} = g^{\alpha\mu} p_\mu, \quad \dot{p}_\alpha = \{p_\alpha, H\} = \Gamma^{\mu\nu}{}_\alpha p_\mu p_\nu.$$

The complete integrability of this system can be shown with the help of a pair of matrices \mathbf{L} and \mathbf{A} with entries defined on the phase space (the cotangent bundle) and satisfying the Lax pair equation [Lax (1968)]

$$\dot{\mathbf{L}} = \{\mathbf{L}, H\} = [\mathbf{L}, \mathbf{A}]. \quad (3.263)$$

It follows from (3.263) that the quantities $I_k \equiv \frac{1}{k} \mathbf{L}^k$ are all constants of the motion. If in addition they commute with each other $\{I_k, I_j\} = 0$ (*Liouville integrability*) then it is possible to integrate the system completely at least in principle (see e.g., [Arnold (1989)]). The Lax representation (3.263) is not unique. In fact, the Lax pair equation is invariant under a transformation of the form

$$\tilde{\mathbf{L}} = \mathbf{U} \mathbf{L} \mathbf{U}^{-1}, \quad \tilde{\mathbf{A}} = \mathbf{U} \mathbf{A} \mathbf{U}^{-1} - \dot{\mathbf{U}} \mathbf{U}^{-1}. \quad (3.264)$$

We see that \mathbf{L} transforms as a tensor while \mathbf{A} transforms as a connection. As we will see, these statements acquire a more precise meaning in the geometric formulation which we will now describe.

Typically, the Lax matrices are linear in the momenta and in the geometric setting they may also be assumed to be homogeneous. This motivates the introduction of two third rank geometrical objects, $L^\alpha{}_\beta{}^\gamma$ and $A^\alpha{}_\beta{}^\gamma$, such that the Lax matrices can be written in the form [Rosquist (1997)]

$$\mathbf{L} = (L^\alpha{}_\beta) = (L^\alpha{}_\beta{}^\mu p_\mu), \quad \mathbf{A} = (A^\alpha{}_\beta) = (A^\alpha{}_\beta{}^\mu p_\mu).$$

We will refer to $L^\alpha{}_\beta{}^\gamma$ and $A^\alpha{}_\beta{}^\gamma$ as the Lax tensor and the Lax connection, respectively. Defining

$$\mathbf{B} = (B^\alpha{}_\beta) = (B^\alpha{}_\beta{}^\mu p_\mu) = \mathbf{A} - \mathbf{\Gamma}, \quad \text{where} \quad \mathbf{\Gamma} = (\Gamma^\alpha{}_\beta) = (\Gamma^\alpha{}_\beta{}^\mu p_\mu)$$

is the *Levi-Civita connection* with respect to $g_{\alpha\beta}$, it then follows that the Lax pair equation takes the covariant form (see [Rosquist (1997)] for details)

$$L^\alpha{}_\beta{}^{(\gamma;\delta)} = L^\alpha{}_\mu{}^{(\gamma} B^{|\mu|}{}_\beta{}^{\delta)} - B^\alpha{}_\mu{}^{(\gamma} L^{|\mu|}{}_\beta{}^{\delta)},$$

where $L^\alpha{}_\beta{}^\gamma$ and $B^\alpha{}_\beta{}^\gamma$ are tensorial objects. Note that the right-hand side of this equation is traceless, so that upon contracting over α and β we

get the Killing vector equation $L^\mu_{\mu(\alpha;\beta)} = 0$. Splitting the Lax tensors in symmetric and antisymmetric parts with respect to the first two indices,

$$\begin{aligned} S_{\alpha\beta\gamma} &= L_{(\alpha\beta)\gamma}, & R_{\alpha\beta\gamma} &= L_{[\alpha\beta]\gamma} \quad \text{and} \\ P_{\alpha\beta\gamma} &= B_{(\alpha\beta)\gamma}, & Q_{\alpha\beta\gamma} &= B_{[\alpha\beta]\gamma}, \end{aligned}$$

the Lax pair equation can be written as the system [Rosquist and Goliath (1997)]

$$\begin{aligned} S^{\alpha\beta}_{(\gamma;\delta)} &= -2S^{\alpha}_{\mu(\gamma}Q^{\beta)\mu}_{\delta)} + 2R^{\alpha}_{\mu(\gamma}P^{\beta)\mu}_{\delta)}, \\ R^{\alpha\beta}_{(\gamma;\delta)} &= -2R^{\alpha}_{\mu(\gamma}Q^{\beta)\mu}_{\delta)} + 2S^{\alpha}_{\mu(\gamma}P^{\beta)\mu}_{\delta)}. \end{aligned}$$

It is evident that this system is coupled via $P_{\alpha\beta\gamma}$. Setting $P_{\alpha\beta\gamma} = 0$ gives the two separate sets of equations

$$\begin{aligned} S^{\alpha\beta}_{(\gamma;\delta)} &= -2S^{\alpha}_{\mu(\gamma}Q^{\beta)\mu}_{\delta)}, \\ R^{\alpha\beta}_{(\gamma;\delta)} &= -2R^{\alpha}_{\mu(\gamma}Q^{\beta)\mu}_{\delta)}. \end{aligned}$$

We will see below that the Lax tensors $L_{\alpha\beta\gamma}$ and $B_{\alpha\beta\gamma}$ in a geometrized version of the open Toda lattice are symmetric and antisymmetric respectively and therefore satisfy (3.265). If $R_{\alpha\beta\gamma}$ is totally antisymmetric (with respect to all three indices) and $Q_{\alpha\beta\gamma} = 0$, then the equations (3.265) are identical to the third rank Killing–Yano equations [Yano (1952)]. Therefore third rank Killing–Yano tensors are special cases of Lax tensors.

It is possible but not necessary to identify the invariant I_2 with the geodesic Hamiltonian (3.262). If such an identification is done then the metric is given by $g^{\alpha\beta} = L^\mu_{\nu}{}^{\alpha}L^\nu_{\mu}{}^{\beta}$. Defining matrices \mathbf{L}^μ with components $(\mathbf{L}^\mu)^\alpha{}_\beta = L^\alpha{}_\beta{}^\mu$, the metric components are given by $g^{\alpha\beta} = (\mathbf{L}^\alpha\mathbf{L}^\beta)$, which suggests using the components of the \mathbf{L}^μ (or some internal variables from which the \mathbf{L}^μ are built) as the basic variables already in the formulation of the field equations much like in the Ashtekar variable formalism [Ashtekar (1988)].

3.16.2 Geometrization of the 3-Particle Open Toda Lattice

Integrable systems are usually discussed in the context of classical mechanics. Classical Hamiltonians typically consist of a flat positive-definite kinetic energy together with a potential energy term. They are thus superficially quite different from geometric Hamiltonians of the form (3.262). However, any classical Hamiltonian with a quadratic kinetic energy can be transformed to a geometric representation. One such geometrization

results in the Jacobi Hamiltonian [Lanczos (1986)]. Another closely related geometrization was used in [Rosquist (1997)]. Both methods involve a re-parameterization of the independent variable. Usually we will refer to the independent variable as the time, although its physical interpretation may vary. As a consequence of this feature, the original Lax representation is not preserved. It is known how to transform the invariants themselves under the time re-parameterization [Rosquist and Pucacco (1995); Rosquist (1997)]. Given that the geometrized invariants are also in involution, the existence of a Lax representation is guaranteed [Babelon and Viallet (1990)]. However, to actually find such a Lax representation is non-trivial. Another geometrization scheme which does preserve the original Lax representation is to apply a suitable canonical transformation. This is however only possible for Hamiltonians with a potential of a special form. One such system that we will consider in this paper is the 3-particle open Toda lattice

$$H = \frac{1}{2} (\bar{p}_1^2 + \bar{p}_2^2 + \bar{p}_3^2) + e^{2(\bar{q}^1 - \bar{q}^2)} + e^{2(\bar{q}^2 - \bar{q}^3)}. \quad (3.265)$$

Below we will discuss two canonical transformations which correspond to inequivalent geometric representations of (3.265). For an explicit integration of the Toda lattice, see e.g. [Perelomov (1990)]. The standard symmetric Lax representation is [Perelomov (1990)]

$$\mathbf{L} = \begin{pmatrix} \bar{p}_1 & \bar{a}_1 & 0 \\ \bar{a}_1 & \bar{p}_2 & \bar{a}_2 \\ 0 & \bar{a}_2 & \bar{p}_3 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 0 & \bar{a}_1 & 0 \\ -\bar{a}_1 & 0 & \bar{a}_2 \\ 0 & -\bar{a}_2 & 0 \end{pmatrix},$$

where $\bar{a}_1 = \exp(\bar{q}^1 - \bar{q}^2), \quad \bar{a}_2 = \exp(\bar{q}^2 - \bar{q}^3).$

Note that the definitions of \bar{a}_1 and \bar{a}_2 differ from the ones used in [Rosquist (1997)]. The Hamiltonian (3.265) admits the linear invariant, $I_1 = \mathbf{L} = \bar{p}_1 + \bar{p}_2 + \bar{p}_3$, corresponding to translational invariance. The Lax representation also gives rise to the two invariants, $I_2 = \frac{1}{2} \mathbf{L}^2 = H$ and $I_3 = \frac{1}{3} \mathbf{L}^3$. We will assume that the tensorial Lax representation is linear and homogeneous in the momenta. A homogeneous Lax representation can be obtained from the standard representation by applying a canonical transformation of the phase space.

3.16.2.1 Tensorial Lax Representation

Here we straightforwardly apply a simple canonical transformation that will give a linear and homogeneous Lax representation [Rosquist and Goliath

(1997)]

$$\begin{aligned}\bar{q}^1 &= q^1 + \ln p_1, & \bar{p}_1 &= p_1, \\ \bar{q}^2 &= q^2, & \bar{p}_2 &= p_2, \\ \bar{q}^3 &= q^3 - \ln p_3, & \bar{p}_3 &= p_3.\end{aligned}$$

The resulting Lax pair matrices are

$$\mathbf{L} = \begin{pmatrix} p_1 & a_1 p_1 & 0 \\ a_1 p_1 & p_2 & a_2 p_3 \\ 0 & a_2 p_3 & p_3 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 0 & a_1 p_1 & 0 \\ -a_1 p_1 & 0 & a_2 p_3 \\ 0 & -a_2 p_3 & 0 \end{pmatrix},$$

where $a_1 = \exp(q^1 - q^2), \quad a_2 = \exp(q^2 - q^3).$

The Hamiltonian is now purely kinetic

$$H = \frac{1}{2} \mathbf{L}^2 = \frac{1}{2} [(1 + 2a_1^2) p_1^2 + p_2^2 + (1 + 2a_2^2) p_3^2].$$

Using (3.262) we identify a metric

$$ds^2 = g_{11}(dq^1)^2 + (dq^2)^2 + g_{33}(dq^3)^2, \quad \text{where} \\ g_{11} = (1 + 2a_1^2)^{-1}, \quad g_{33} = (1 + 2a_2^2)^{-1}.$$

The non-zero Levi-Civita connection coefficients, $\Gamma_{\beta\gamma}^\alpha = \Gamma_{(\beta\gamma)}^\alpha$, of this metric are

$$\begin{aligned}\Gamma_{11}^1 &= -2a_1^2 g_{11}, & \Gamma_{33}^2 &= 2a_2^2 (g_{33})^2, \\ \Gamma_{12}^1 &= 2a_1^2 g_{11}, & \Gamma_{23}^3 &= -2a_2^2 g_{33}, \\ \Gamma_{11}^2 &= -2a_1^2 (g_{11})^2, & \Gamma_{33}^3 &= 2a_2^2 g_{33}.\end{aligned}$$

Following the arguments above, the homogeneous Lax matrix should correspond to a tensor with mixed indices L_{β}^α . It is a reasonable assumption that the covariant Lax formulation inherits the symmetries of the standard formulation we started with. We therefore expect $L_{\alpha\beta}$ and $B_{\alpha\beta}$ to have the symmetries

$$L_{(\alpha\beta)} = L_{\alpha\beta} \quad \text{and} \quad B_{[\alpha\beta]} = B_{\alpha\beta}.$$

Note that the symmetry properties are not imposed on the Lax matrices, L_{β}^α and B_{β}^α , themselves. In fact, the required symmetries are not consistent with the representation (3.266). We can however perform a similarity transformation (3.264) of the Lax matrix, $\mathbf{L} \rightarrow \tilde{\mathbf{L}}$ in such a way that $\tilde{L}_{\alpha\beta}$

will be symmetric. Using the transformation matrix

$$\mathbf{U} = \begin{pmatrix} 1/\sqrt{g_{11}} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/\sqrt{g_{33}} \end{pmatrix},$$

will get a new Lax pair

$$\mathbf{L} = \begin{pmatrix} p_1 & a_1/\sqrt{g_{11}} p_1 & 0 \\ a_1\sqrt{g_{11}} p_1 & p_2 & a_2\sqrt{g_{33}} p_3 \\ 0 & a_2/\sqrt{g_{33}} p_3 & p_3 \end{pmatrix},$$

$$\mathbf{A} = \begin{pmatrix} \Gamma_1^{11} p_1 + \Gamma_1^{12} p_2 & a_1/\sqrt{g_{11}} p_1 & 0 \\ -a_1\sqrt{g_{11}} p_1 & 0 & a_2\sqrt{g_{33}} p_3 \\ 0 & -a_2/\sqrt{g_{33}} p_3 & \Gamma_3^{32} p_2 + \Gamma_3^{33} p_3 \end{pmatrix},$$

where \mathbf{L} is such that $L_{\alpha\beta}$ is symmetric. Defining $\hat{\mathbf{L}} = (L_{\alpha\beta})$ and $\hat{\mathbf{A}} = (A_{\alpha\beta})$ by

$$\hat{\mathbf{L}} = \mathbf{g}\mathbf{L}, \quad \hat{\mathbf{A}} = \mathbf{g}\mathbf{A}, \quad \text{where} \quad \mathbf{g} = (g_{\alpha\beta}), \quad \text{we get}$$

$$\hat{\mathbf{L}} = \begin{pmatrix} g_{11}p_1 & a_1\sqrt{g_{11}} p_1 & 0 \\ a_1\sqrt{g_{11}} p_1 & p_2 & a_2\sqrt{g_{33}} p_3 \\ 0 & a_2\sqrt{g_{33}} p_3 & g_{33}p_3 \end{pmatrix}, \quad \text{and}$$

$$\hat{\mathbf{A}} = \begin{pmatrix} \Gamma_{11}^1 p_1 + \Gamma_{11}^2 p_2 & a_1\sqrt{g_{11}} p_1 & 0 \\ -a_1\sqrt{g_{11}} p_1 & 0 & a_2\sqrt{g_{33}} p_3 \\ 0 & -a_2\sqrt{g_{33}} p_3 & \Gamma_{33}^2 p_2 + \Gamma_{33}^3 p_3 \end{pmatrix}.$$

Note that the upper triangular parts of $\hat{\mathbf{L}}$ and $\hat{\mathbf{A}}$ coincide. This property is peculiar to the open Toda lattice. We also define the corresponding *connection matrix*, $\hat{\mathbf{\Gamma}} = \mathbf{g}\mathbf{\Gamma}$, given by

$$\hat{\mathbf{\Gamma}} = \begin{pmatrix} \Gamma_{11}^1 p_1 + \Gamma_{11}^2 p_2 & 2a_1^2 g_{11} p_1 & 0 \\ -2a_1^2 g_{11} p_1 & 0 & 2a_2^2 g_{33} p_3 \\ 0 & -2a_2^2 g_{33} p_3 & \Gamma_{33}^2 p_2 + \Gamma_{33}^3 p_3 \end{pmatrix}.$$

We see that the off-diagonal part of the matrix $\hat{\mathbf{\Gamma}}$ is antisymmetric like that of $\hat{\mathbf{A}}$ and furthermore that their off-diagonal components are related by the simple relation

$$\Gamma_{\alpha\beta}^\gamma = 2(A_{\alpha\beta}^\gamma)^2, \quad (\text{for } \alpha < \beta).$$

This gives the following relation between the upper triangular parts of $\hat{\mathbf{L}}$ and $\hat{\mathbf{A}}$

$$\Gamma_{\alpha\beta}{}^\gamma = 2(L_{\alpha\beta}{}^\gamma)^2, \quad (\text{for } \alpha < \beta).$$

Using the relation $\hat{\mathbf{A}} = \hat{\mathbf{\Gamma}} + \hat{\mathbf{B}}$, where $\hat{\mathbf{B}} = \mathbf{gB}$, we find the following relation between the upper triangular components of $\hat{\mathbf{L}}$ and $\hat{\mathbf{B}}$

$$B_{\alpha\beta}{}^\gamma = L_{\alpha\beta}{}^\gamma - 2(L_{\alpha\beta}{}^\gamma)^2, \quad (\text{for } \alpha < \beta).$$

Finally expressing $\hat{\mathbf{L}}$ and $\hat{\mathbf{B}}$ in terms of $\hat{\mathbf{\Gamma}}$ we have for the upper triangular parts

$$L_{\alpha\beta}{}^\gamma = \sqrt{\frac{1}{2}\Gamma_{\alpha\beta}{}^\gamma}, \quad B_{\alpha\beta}{}^\gamma = -\Gamma_{\alpha\beta}{}^\gamma + \sqrt{\frac{1}{2}\Gamma_{\alpha\beta}{}^\gamma}, \quad (\text{for } \alpha < \beta).$$

Furthermore, the diagonal elements of $\hat{\mathbf{A}}$ and $\hat{\mathbf{\Gamma}}$ are identical. This implies that $\hat{\mathbf{B}}$ is antisymmetric in agreement with our expectations.

3.16.3 4D Generalizations

Recall that we can get a 4D space–time simply by adding a time coordinate according to the prescription [Rosquist and Goliath (1997)]

$${}^{(4)}ds^2 = -(dq^0)^2 + ds^2,$$

where ds^2 is a 3D positive–definite metric. It follows that ${}^{(4)}\mathbf{\Gamma} = \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{\Gamma} \end{pmatrix}$.

For the cases obtained above this will lead to inequivalent space–times. One way to generalize the 3D Lax pair is

$$\begin{aligned} {}^{(4)}\mathbf{L} &= \begin{pmatrix} i p_0 & 0 \\ 0 & \mathbf{L} \end{pmatrix}, & {}^{(4)}\mathbf{A} &= \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{A} \end{pmatrix}, \\ \text{for which } {}^{(4)}\mathbf{B} &= \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{B} \end{pmatrix}. \end{aligned}$$

This Lax pair gives the geodesic Hamiltonian of the corresponding space–time metric as quadratic invariant.

3.16.3.1 Case I

Adding a time dimension to (3.266) we get the metric [Rosquist and Goliath (1997)]

$$ds^2 = -(dq^0)^2 + g_{11}(dq^1)^2 + (dq^2)^2 + g_{33}(dq^3)^2, \quad \text{where}$$

$$g_{11} = (1 + 2a_1^2)^{-1}, \quad g_{33} = (1 + 2a_2^2)^{-1}, \quad \text{and} \\ a_1 = \exp(q^1 - q^2), \quad a_2 = \exp(q^2 - q^3).$$

This space-time is of Petrov type I. The nonzero components of the energy-momentum tensor calculated in a Lorentz frame are ($\kappa = 1$)

$$\begin{aligned} T^{00} &= -((g_{11})^2 T^{11} + T^{22} + (g_{33})^2 T^{33}) \\ &= -4a_1^2 (g_{11})^2 (a_1^2 - 1) + 4a_1^2 a_2^2 g_{11} g_{33} - 4a_2^2 (g_{33})^2 (a_2^2 - 1), \\ T^{11} &= 4a_1^2 (a_1^2 - 1), \quad T^{22} = -4a_1^2 a_2^2 g_{11} g_{33}, \\ T^{33} &= 4a_2^2 (a_2^2 - 1). \end{aligned}$$

3.16.3.2 Case II

Here we have the metric [Rosquist and Goliath (1997)]

$$\begin{aligned} ds^2 &= -(dq^0)^2 + g_{11}(dq^1)^2 + (dq^2)^2 + (dq^3)^2, \quad \text{where} \\ g_{11} &= [1 + 2(a_1^2 + a_2^2)]^{-1}, \quad a_1 = \exp\left(\frac{1}{\sqrt{2}}q^1 + \sqrt{\frac{3}{2}}q^2\right), \\ a_2 &= \exp\left(\frac{1}{\sqrt{2}}q^1 - \sqrt{\frac{3}{2}}q^2\right). \end{aligned}$$

This space-time is of Petrov type D. The nonzero components of the energy-momentum tensor calculated in a Lorentz frame are ($\kappa = 1$)

$$\begin{aligned} T^{00} &= 12 e^{2\sqrt{2}q^1} g_{11}^2 \left(4 - 2 \sinh^2(\sqrt{6}q^2) + e^{-\sqrt{2}q^1} \cosh(\sqrt{6}q^2)\right), \\ T^{11} &= -(g^{11})^2 T^{00}. \end{aligned}$$

3.16.3.3 Energy-Momentum Tensors

In both of the above cases, the energy-momentum tensor takes the form [Rosquist and Goliath (1997)]

$$T^{\alpha\beta} = \begin{pmatrix} \mu & 0 & 0 & 0 \\ 0 & p_1 & 0 & 0 \\ 0 & 0 & p_2 & 0 \\ 0 & 0 & 0 & p_3 \end{pmatrix},$$

where $\mu \equiv T^{00}$ is the energy density, and $p_i \equiv T^{ii}$, ($i = 1, 2, 3$) are anisotropic pressures. Such an energy-momentum tensor is physically meaningful if the *weak energy condition* [Hawking and Ellis (1973)]

$$\mu \geq 0, \quad \mu + p_i \geq 0, \quad (i = 1, 2, 3),$$

is satisfied. For case I, there is an unbounded sub-domain of the space coordinates (q^1, q^2, q^3) for which the weak energy condition holds. For case II, it is easily seen that the restrictions on the energy-momentum tensor are inconsistent, so that the weak energy condition never holds.

3.17 Applied Unorthodox Geometries

3.17.1 Noncommutative Geometry

In this subsection we give review of noncommutative geometry and its main gravitational applications. In the last section of the book, we will give its applications to string theory.

3.17.1.1 Moyal Product and Noncommutative Algebra

Noncommutative geometry is concerned with the possible spatial interpretations of algebraic structures for which the commutative law fails; that is, for which xy does not always equal yx . The challenge of the theory is to get around the lack of commutative multiplication, which is a requirement of previous geometric theories of such structures (see [Connes (1994)]).

Recall that an ordinary differentiable manifold can be characterized by the commutative algebra of smooth functions defined on it, and the space of smooth sections of its tangent bundle, cotangent bundle and other fiber bundles. All these spaces are modules over the commutative algebra of smooth functions. The concepts of exterior derivative, Lie derivative and covariant derivative are also important elements in understanding derivations over this algebra. In the noncommutative case, the algebras in question are noncommutative. To handle differential forms, one must work with the graded exterior algebra bundle of all p -forms under the wedge product and look at its algebra of smooth sections. A ‘differential’ is taken to be an anti-derivation (or, something more general) on this algebra, which increases the grading by 1 and is quadratically nilpotent.

Historically first noncommutative product was the *Moyal product* [Moyal (1949)],¹⁰ that is an associative, noncommutative ‘star’-product. For any two functions f, g on a Poisson manifold M , the Moyal product $*$

¹⁰The Moyal product is also sometimes called Weyl-Moyal product.

is defined as:

$$f * g = fg + \sum_{n=1}^{\infty} \hbar^n C_n(f, g),$$

where each C_n is a bi-differential operator of order n with the following properties:

1. $f * g = fg + O(\hbar)$ – deformation of the pointwise product.
2. $f * g - g * f = i\{f, g\} + O(\hbar^2)$ – deformation in the direction of the Poisson bracket.
3. $f * 1 = 1 * f = f$ – the 1 of the un-deformed algebra is the 1 in the new algebra.
4. $\overline{f * g} = \overline{g} * \overline{f}$ – the complex conjugate is an anti-linear anti-automorphism.

Let the *Poisson structure* on a symplectic manifold M be defined by

$$\omega = \omega^{ij} \partial_i \wedge \partial_j.$$

If this structure is constant, that is, if ω^{ij} do not depend on the local coordinates on M^{11} , then the Moyal product of two functions $f, g \in M$ can be defined as

$$f * g = fg + \hbar \omega^{ij} (\partial_i f) (\partial_j g) + \frac{\hbar^2}{2} \omega^{ij} \omega^{km} (\partial_i \partial_k f) (\partial_j \partial_m g) + \dots$$

where \hbar is the (reduced) *Planck constant*.

For example, in *Weyl deformation quantization* [Weyl (1927)], the symplectic phase-space of classical mechanics is deformed into a *noncommutative phase-space* generated by the position and momentum operators, using the Moyal product.

3.17.1.2 Noncommutative Space-Time Manifolds

In physical field theories one usually considers differential space-time manifolds. Now, in the noncommutative realm, the notion of a point is no longer well-defined and we have to give up the concept of differentiable manifolds. However, the space of functions on a manifold forms an algebra. A generalization of this algebra can be considered in the noncommutative case. We take the algebra freely generated by the *noncommutative coordinates* $\{\hat{x}^i\}$, which respects commutation relations of the type [Madore (1995)]

$$[\hat{x}^\mu, \hat{x}^\nu] = C^{\mu\nu}(\hat{x}) \neq 0. \quad (3.266)$$

¹¹Such a form can always be found at least locally by *Darboux's Theorem*

We can take the space of formal power series in the coordinates \hat{x}^i and divide by the ideal generated by the above relations

$$\hat{\mathcal{A}}_{\hat{x}} = \mathbb{C}\langle\langle\hat{x}^0, \dots, \hat{x}^n\rangle\rangle / ([\hat{x}^\mu, \hat{x}^\nu] - C^{\mu\nu}(\hat{x})),$$

where the function $C^{\mu\nu}(\hat{x})$ is unknown. For physical reasons, $C^{\mu\nu}(\hat{x})$ should be a function that vanishes at large distances where we experience the commutative world and may be determined by experiments. Nevertheless, one can consider its power-series expansion [Meyer (2005)]

$$C^{\mu\nu}(\hat{x}) = i\theta^{\mu\nu} + iC^{\mu\nu}{}_{\rho}\hat{x}^{\rho} + (q\hat{R}^{\mu\nu}{}_{\rho\sigma} - \delta_{\rho}^{\nu}\delta_{\sigma}^{\mu})\hat{x}^{\rho}\hat{x}^{\sigma} + \dots,$$

where $\theta^{\mu\nu}$, $C^{\mu\nu}{}_{\rho}$ and $q\hat{R}^{\mu\nu}{}_{\rho\sigma}$ are constants, and study cases where the commutation relations are constant, linear or quadratic in the \hat{x}^i -coordinates. At very short distances, these cases provide a reasonable approximation for $C^{\mu\nu}(\hat{x})$ and lead to the following three structures:

(1) Canonical, or θ -deformed case:

$$[\hat{x}^\mu, \hat{x}^\nu] = i\theta^{\mu\nu}. \quad (3.267)$$

(2) Lie-algebra case:

$$[\hat{x}^\mu, \hat{x}^\nu] = iC^{\mu\nu}{}_{\rho}\hat{x}^{\rho}. \quad (3.268)$$

(3) Quantum spaces case:

$$\hat{x}^\mu\hat{x}^\nu = q\hat{R}^{\mu\nu}{}_{\rho\sigma}\hat{x}^{\rho}\hat{x}^{\sigma}. \quad (3.269)$$

We denote by $\hat{\mathcal{A}}$ the algebra generated by noncommutative coordinates \hat{x}^μ which are subject to the relations (3.267) and call it the *algebra of noncommutative functions*. The algebra of the corresponding commutative functions will be denoted by \mathcal{A} . We consider the θ -deformed case (3.267), but we note that the algebraic construction presented here can be generalized to more complicated noncommutative structures of the above type, which possess the so-called *Poincaré-Birkhoff-Witt property*, which states that the space of polynomials in noncommutative coordinates of a given degree is isomorphic to the space of polynomials in the commutative coordinates. Such an isomorphism between polynomials of a fixed degree is given by an ordering prescription. One example is the *symmetric Weyl ordering*, denoted W , which assigns to any monomial the totally symmetric

ordered monomial [Meyer (2005)]

$$W : \mathcal{A} \rightarrow \hat{\mathcal{A}}, \quad x^\mu \mapsto \hat{x}^\mu, \quad x^\mu x^\nu \mapsto \frac{1}{2}(\hat{x}^\mu \hat{x}^\nu + \hat{x}^\nu \hat{x}^\mu) \cdots \quad (3.270)$$

To study the dynamics of physical fields we need a differential calculus on the noncommutative algebra $\hat{\mathcal{A}}$. Derivatives are maps on the deformed coordinate space [Wess (2004)]

$$\hat{\partial}_\mu : \hat{\mathcal{A}} \rightarrow \hat{\mathcal{A}}.$$

This means that they have to be consistent with the commutation relations of the coordinates. In the θ -constant case a consistent differential calculus can be defined very easily by¹²

$$[\hat{\partial}_\mu, \hat{x}^\nu] = \delta_\mu^\nu, \quad (\hat{\partial}_\mu, \hat{\partial}_\nu) = 0. \quad (3.271)$$

It is the fully undeformed differential calculus. The above definitions yield the usual Leibniz-rule for the derivatives $\hat{\partial}_\mu$

$$(\hat{\partial}_\mu \hat{f} \hat{g}) = (\hat{\partial}_\mu \hat{f}) \hat{g} + \hat{f} (\hat{\partial}_\mu \hat{g}). \quad (3.272)$$

This is a special feature of the fact that $\theta^{\mu\nu}$ are constants. In the more complicated examples of noncommutative structures this undeformed Leibniz-rule usually cannot be preserved but one has to consider deformed Leibniz-rules for the derivatives [Wess and Zumino (1991)]. Note that (3.271) also implies that

$$(\hat{\partial}_\mu \hat{f}) = \widehat{(\partial_\mu f)}. \quad (3.273)$$

The Weyl ordering (3.270) can be formally implemented by the map

$$f \mapsto W(f) = \frac{1}{(2\pi)^{\frac{n}{2}}} \int d^n k e^{ik_\mu \hat{x}^\mu} \tilde{f}(k)$$

where \tilde{f} is the *Fourier transform* of f

$$\tilde{f}(k) = \frac{1}{(2\pi)^{\frac{n}{2}}} \int d^n x e^{-ik_\mu x^\mu} f(x).$$

This is due to the fact that the exponential is a fully symmetric function. Using the *Baker–Campbell–Hausdorff formula* one finds

$$e^{ik_\mu \hat{x}^\mu} e^{ip_\nu \hat{x}^\nu} = e^{i(k_\mu + p_\mu) \hat{x}^\mu - \frac{1}{2} k_\mu \theta^{\mu\nu} p_\nu}. \quad (3.274)$$

¹²We use brackets to distinguish the action of a differential operator from the multiplication in the algebra of differential operators.

This immediately leads to the following observation

$$\begin{aligned}\hat{f}\hat{g} &= W(f)W(g) = \frac{1}{(2\pi)^n} \int d^n k d^n p e^{ik_\mu \hat{x}^\mu} e^{ip_\nu \hat{x}^\nu} \tilde{f}(k) \tilde{g}(p) = \\ &\frac{1}{(2\pi)^n} \int d^n k d^n p e^{i(k_\mu + p_\mu) \hat{x}^\mu} e^{-\frac{i}{2} k_\mu \theta^{\mu\nu} p_\nu} \tilde{f}(k) \tilde{g}(p) = W(\mu \circ e^{\frac{i}{2} \theta^{\mu\nu} \partial_\mu \otimes \partial_\nu} f \otimes g),\end{aligned}\quad (3.275)$$

where $\mu(f \otimes g) := fg$ is the multiplication map. With (3.273) we deduce from (3.275) the equation

$$\mu \circ e^{-\frac{i}{2} \theta^{\mu\nu} \hat{\partial}_\mu \otimes \hat{\partial}_\nu} \hat{f} \otimes \hat{g} = \widehat{fg}. \quad (3.276)$$

The above formula shows us how the commutative and the noncommutative product are related. It will be important for us later on.

3.17.1.3 Symmetries and Diffeomorphisms on Deformed Spaces

In general, the commutation relations (3.266) are *not* covariant with respect to undeformed symmetries. For example, the canonical commutation relations (3.267) break Lorentz symmetry if we assume that the noncommutativity parameters $\theta^{\mu\nu}$ do not transform.

The question arises whether we can *deform* the symmetry in such a way that it acts consistently on the deformed space (i.e., leaves the deformed space invariant) and such that it reduces to the undeformed symmetry in the commutative limit [Meyer (2005)]. The answer is yes: Lie algebras can be deformed in the *category of Hopf algebras*.¹³ Important examples of such deformations are q -deformations: there exists a q -deformation of the universal enveloping algebra of an arbitrary semisimple Lie algebra.¹⁴ Module algebras of this q -deformed universal enveloping algebras are noncommutative spaces with commutation relations of type (3.269). There exists also a so-called κ -deformation of the *Poincaré algebra* (see [Dimitrijevic *et al.* (2003); Dimitrijevic *et al.* (2004)]), which leads to a noncommutative space of the Lie type (3.268).

Quantum group symmetries lead to new features of field theories on noncommutative spaces. Because of its simplicity, θ -deformed spaces are very well-suited to study them. For results on the consequences of the θ -deformed Poincaré algebra, see [Chaichian *et al.* (2004); Aschieri *et al.* (2005)].

¹³Hopf algebras coming from a Lie algebra are also called Quantum groups.

¹⁴It is called q -deformation since it is a deformation in terms of a parameter q .

Now, recall that gravity is a theory invariant with respect to diffeomorphisms. However, to generalize the Einstein formalism to noncommutative spaces in order to establish a *noncommutative gravity theory*, it is important to first understand that diffeomorphisms possess more geometrical structure than the algebraic one: They are naturally equipped with a *Hopf algebra* structure.

Recall that diffeomorphisms are generated by vector-fields ξ . Acting on functions, vector-fields are represented as linear differential operators $\xi = \xi^\mu \partial_\mu$. Vector-fields form a Lie algebra Ξ over the field \mathbb{C} with the *Lie bracket* given by

$$[\xi, \eta] = \xi \times \eta,$$

where $\xi \times \eta$ is defined by its action on functions

$$(\xi \times \eta)(f) = (\xi^\mu (\partial_\mu \eta^\nu) \partial_\nu - \eta^\mu (\partial_\mu \xi^\nu) \partial_\nu)(f).$$

The Lie algebra of *infinitesimal diffeomorphisms* Ξ can be embedded into its universal enveloping algebra which we want to denote by $\mathcal{U}(\Xi)$. The universal enveloping algebra is an associative algebra and possesses a natural Hopf algebra structure. It is given by the following structure maps [Meyer (2005)]:

- An algebra homomorphism called *coproduct* defined by

$$\Delta : \mathcal{U}(\Xi) \rightarrow \mathcal{U}(\Xi) \otimes \mathcal{U}(\Xi), \quad \Xi \ni \xi \mapsto \Delta(\xi) := \xi \otimes 1 + 1 \otimes \xi. \quad (3.277)$$

- An algebra homomorphism called *counit* defined by

$$\varepsilon : \mathcal{U}(\Xi) \rightarrow \mathbb{C}, \quad \Xi \ni \xi \mapsto \varepsilon(\xi) = 0. \quad (3.278)$$

- An anti-algebra homomorphism called *antipode* defined by

$$S : \mathcal{U}(\Xi) \rightarrow \mathcal{U}(\Xi), \quad \Xi \ni \xi \mapsto S(\xi) = -\xi. \quad (3.279)$$

For a precise definition and more details on Hopf algebras we refer the reader to the text-books [Chari and Presley (1995); Klimyk and Schmüdgen (1997)]. For our purposes it shall be sufficient to note that the coproduct implements how the Hopf algebra acts on a product in a representation algebra. It is now possible to study deformations of $\mathcal{U}(\Xi)$ in the category of Hopf algebras. This leads to a deformed version of diffeomorphisms, which is the fundamental building block of our approach to noncommutative gravity theory.

Recall that scalar fields are defined by their transformation property with respect to infinitesimal coordinate transformations:

$$\delta_\xi \phi = -\xi^\mu \phi = -\xi^\mu (\partial_\mu \phi). \quad (3.280)$$

The product of two scalar fields is transformed using the *Leibniz rule*

$$\delta_\xi(\phi\psi) = (\delta_\xi \phi)\psi + \phi(\delta_\xi \psi) = -\xi^\mu (\partial_\mu \phi\psi), \quad (3.281)$$

such that the product of two scalar fields transforms again as a scalar. The above Leibniz rule can be understood in mathematical terms as follows [Meyer (2005)]: The Hopf algebra $\mathcal{U}(\Xi)$ is represented on the space of scalar fields by infinitesimal coordinate transformations δ_ξ . On scalar fields, the action of δ_ξ is explicitly given by the differential operator $-\xi^\mu \partial_\mu$. Recall that the space of scalar fields is not only a vector space, as it possesses also an algebra structure, such as $\mathcal{U}(\Xi)$ is not only an algebra, but also a Hopf algebra, as it possesses in addition the co-structure maps defined above. We say that a Hopf algebra H acts on an algebra A if A is a module with respect to the algebra H and if in addition for all $h \in H$ and $a, b \in A$

$$h(ab) = \mu \circ \Delta h(a \otimes b), \quad h(1) = \varepsilon(h).$$

Here μ is the multiplication map defined by $\mu(a \otimes b) = ab$. In our concrete example where $H = \mathcal{U}(\Xi)$ and A is the algebra of scalar fields we indeed have that the algebra of scalar fields is a $\mathcal{U}(\Xi)$ -module algebra. This can be seen easily if we rewrite (3.281) using (3.277) for the generators $\xi \in \Xi$ for $\mathcal{U}(\Xi)$:

$$\delta_\xi(\phi\psi) = (\delta_\xi \phi)\psi + \phi(\delta_\xi \psi) = \mu \circ \Delta \xi(\phi \otimes \psi).$$

It is also evident that

$$\delta_\xi 1 = 0 = \varepsilon(\xi)1.$$

Now we are in the right mathematical framework: We study a Lie algebra (here infinitesimal diffeomorphisms Ξ) and embed it in its universal enveloping algebra (here $\mathcal{U}(\Xi)$). This universal enveloping algebra is a Hopf algebra via a natural Hopf structure induced by (3.277, 3.278, 3.279).

Physical quantities live in representations of this Hopf algebras. For instance, the algebra of scalar fields is a $\mathcal{U}(\Xi)$ -module algebra. The action of $\mathcal{U}(\Xi)$ on scalar fields is given in terms of infinitesimal coordinate transformations δ_ξ .

Similarly one studies tensor representations of $\mathcal{U}(\Xi)$. For example vector fields are introduced by the transformation property

$$\delta_\xi V_\alpha = -\xi^\mu (\partial_\mu V_\alpha) - (\partial_\alpha \xi^\mu) V_\mu, \quad \delta_\xi V^\alpha = -\xi^\mu (\partial_\mu V^\alpha) + (\partial_\mu \xi^\alpha) V^\mu.$$

The generalization to arbitrary tensor fields is straight forward:

$$\begin{aligned} \delta_\xi T_{\nu_1 \dots \nu_n}^{\mu_1 \dots \mu_n} = & -\xi^\mu (\partial_\mu T_{\nu_1 \dots \nu_n}^{\mu_1 \dots \mu_n}) + (\partial_\mu \xi^{\mu_1}) T_{\nu_1 \dots \nu_n}^{\mu_1 \dots \mu_n} + \dots + (\partial_\mu \xi^{\mu_n}) T_{\nu_1 \dots \nu_n}^{\mu_1 \dots \mu_n} \\ & - (\partial_{\nu_1} \xi^\nu) T_{\nu_1 \dots \nu_n}^{\mu_1 \dots \mu_n} - \dots - (\partial_{\nu_n} \xi^\nu) T_{\nu_1 \dots \nu_n}^{\mu_1 \dots \mu_n}. \end{aligned}$$

As for scalar fields, we also find that the product of two tensors transforms like a tensor. Summarizing, we have seen that scalar fields, vector fields and tensor fields are representations of the Hopf algebra $\mathcal{U}(\Xi)$, the universal enveloping algebra of infinitesimal diffeomorphisms. The Hopf algebra $\mathcal{U}(\Xi)$ acts via *infinitesimal coordinate transformations* δ_ξ which are subject to the relations:

$$[\delta_\xi, \delta_\eta] = \delta_{\xi \times \eta} \varepsilon(\delta_\xi) = 0, \quad \Delta \delta_\xi = \delta_\xi \otimes 1 + 1 \otimes \delta_\xi S(\delta_\xi) = -\delta_\xi. \quad (3.282)$$

The transformation operator δ_ξ is explicitly given by differential operators which depend on the representation under consideration. In case of scalar fields this differential operator is given by $-\xi^\mu \partial_\mu$.

3.17.1.4 Deformed Diffeomorphisms

The above concepts can be deformed in order to establish a consistent tensor calculus on the noncommutative space–time algebra (3.267). In this context it is necessary to account the full Hopf algebra structure of the universal enveloping algebra $\mathcal{U}(\Xi)$.

In our setting the algebra $\hat{\mathcal{A}}$ possesses a noncommutative product defined by

$$[\hat{x}^\mu, \hat{x}^\nu] = i\theta^{\mu\nu}. \quad (3.283)$$

We want to deform the structure maps (3.282) of the Hopf algebra $\mathcal{U}(\Xi)$ in such a way that the resulting deformed Hopf algebra which we denote by $\mathcal{U}(\hat{\Xi})$ consistently acts on $\hat{\mathcal{A}}$. In the language introduced in the previous section this means that we want $\hat{\mathcal{A}}$ to be a $\mathcal{U}(\hat{\Xi})$ –module algebra. We claim that the following deformation of $\mathcal{U}(\Xi)$ does the job. Let $\mathcal{U}(\hat{\Xi})$ be generated as algebra by elements $\hat{\delta}_\xi$, $\xi \in \Xi$. We leave the algebra relation undeformed and demand [Meyer (2005)]

$$[\hat{\delta}_\xi, \hat{\delta}_\eta] = \hat{\delta}_{\xi \times \eta} \quad (3.284)$$

but we deform the co-sector

$$\Delta \hat{\delta}_\xi = e^{-\frac{1}{2} h \theta^{\rho\sigma} \hat{\partial}_\rho \otimes \hat{\partial}_\sigma} (\hat{\delta}_\xi \otimes 1 + 1 \otimes \hat{\delta}_\xi) e^{\frac{1}{2} h \theta^{\rho\sigma} \hat{\partial}_\rho \otimes \hat{\partial}_\sigma}, \quad (3.285)$$

where $[\hat{\partial}_\rho, \hat{\delta}_\xi] = \hat{\delta}_{(\partial_\rho \xi)}$.

The deformed coproduct (3.285) reduces to the undeformed one (3.282) in the limit $\theta \rightarrow 0$. Antipode and counit remain undeformed

$$S(\hat{\delta}_\xi) = -\hat{\delta}_\xi \varepsilon(\hat{\delta}_\xi) = 0.$$

We have to check whether the above deformation is a good one in the sense that it leads to a consistent action on $\hat{\mathcal{A}}$. First we need a differential operator acting on fields in $\hat{\mathcal{A}}$ which represents the algebra (3.284). Let us consider the differential operator

$$\hat{X}_\xi := \sum_{n=0}^{\infty} \frac{1}{n!} \left(-\frac{i}{2}\right)^n \theta^{\rho_1 \sigma_1} \dots \theta^{\rho_n \sigma_n} (\hat{\partial}_{\rho_1} \dots \hat{\partial}_{\rho_n} \hat{\xi}^\mu) \hat{\partial}_\mu \hat{\partial}_{\sigma_1} \dots \hat{\partial}_{\sigma_n}. \quad (3.286)$$

This is to be understood like that: A vector-field $\xi = \xi^\mu \partial_\mu$ is determined by its coefficient functions ξ^μ . Before we have seen that there is a vectorspace isomorphism W from the space of commutative to the space of noncommutative functions which is given by the symmetric ordering prescription. The image of a commutative function f under the isomorphism W is denoted by \hat{f}

$$W : f \mapsto W(f) = \hat{f}.$$

In (3.286) $\hat{\xi}^\mu$ is therefore to be interpreted as the image of ξ^μ with respect to W . Then indeed we have

$$[\hat{X}_\xi, \hat{X}_\eta] = \hat{X}_{\xi \times \eta}. \quad (3.287)$$

To see this we use result (3.276) to rewrite $(\hat{X}_\xi \hat{\phi})$:

$$\begin{aligned} (\hat{X}_\xi \hat{\phi}) &= \sum_{n=0}^{\infty} \frac{1}{n!} \left(-\frac{i}{2}\right)^n \theta^{\rho_1 \sigma_1} \dots \theta^{\rho_n \sigma_n} (\hat{\partial}_{\rho_1} \dots \hat{\partial}_{\rho_n} \hat{\xi}^\mu) (\hat{\partial}_\mu \hat{\partial}_{\sigma_1} \dots \hat{\partial}_{\sigma_n} \hat{\phi}) = \quad (3.288) \\ \sum_{n=0}^{\infty} \frac{1}{n!} \left(-\frac{i}{2}\right)^n \theta^{\rho_1 \sigma_1} \dots \theta^{\rho_n \sigma_n} (\hat{\partial}_{\rho_1} \dots \hat{\partial}_{\rho_n} \hat{\xi}^\mu) (\hat{\partial}_{\sigma_1} \dots \hat{\partial}_{\sigma_n} \widehat{\partial_\mu \phi}) &= \widehat{\xi^\mu (\partial_\mu \phi)} = \widehat{(\xi \phi)}. \end{aligned}$$

From (3.288) it follows

$$(\hat{X}_\xi (\hat{X}_\eta \hat{\phi})) - (\hat{X}_\eta (\hat{X}_\xi \hat{\phi})) = (\widehat{[\xi, \eta] \phi}) = (\hat{X}_{\xi \times \eta} \hat{\phi}),$$

which amounts to (3.287). It is therefore reasonable to introduce scalar fields $\hat{\phi} \in \hat{\mathcal{A}}$ by the transformation property

$$\hat{\delta}_\xi \hat{\phi} = -(\hat{X}_\xi \hat{\phi}).$$

The next step is to work out the action of the differential operators \hat{X}_ξ on the product of two fields. A calculation [Aschieri *et. al.* (2005)] shows that

$$(\hat{X}_\xi(\hat{\phi}\hat{\psi})) = \mu \circ (e^{-\frac{1}{2}h\theta^{\rho\sigma}\hat{\partial}_\rho \otimes \hat{\partial}_\sigma}(\hat{X}_\xi \otimes 1 + 1 \otimes \hat{X}_\xi)e^{\frac{1}{2}h\theta^{\rho\sigma}\hat{\partial}_\rho \otimes \hat{\partial}_\sigma} \hat{\phi} \otimes \hat{\psi}).$$

This means that the differential operators \hat{X}_ξ act via a *deformed Leibniz rule* on the product of two fields. Comparing with (3.285) we see that the deformed Leibniz rule of the differential operator \hat{X}_ξ is exactly the one induced by the deformed coproduct (3.285):

$$\hat{\delta}_\xi(\hat{\phi}\hat{\psi}) = e^{-\frac{1}{2}h\theta^{\rho\sigma}\hat{\partial}_\rho \otimes \hat{\partial}_\sigma}(\hat{\delta}_\xi \otimes 1 + 1 \otimes \hat{\delta}_\xi)e^{\frac{1}{2}h\theta^{\rho\sigma}\hat{\partial}_\rho \otimes \hat{\partial}_\sigma}(\hat{\phi}\hat{\psi}) = -\hat{X}_\xi \triangleright (\hat{\phi}\hat{\psi}).$$

Hence, the deformed Hopf algebra $\mathcal{U}(\hat{\Xi})$ is indeed represented on scalar fields $\hat{\phi} \in \hat{\mathcal{A}}$ by the differential operator \hat{X}_ξ . The scalar fields form a $\mathcal{U}(\hat{\Xi})$ -module algebra.

In analogy to the previous section we can introduce vector and tensor fields as representations of the Hopf algebra $\mathcal{U}(\hat{\Xi})$. The transformation property for an arbitrary tensor reads

$$\begin{aligned} \hat{\delta}_\xi \hat{T}_{\nu_1 \dots \nu_s}^{\mu_1 \dots \mu_r} = & -(\hat{X}_\xi \hat{T}_{\nu_1 \dots \nu_n}^{\mu_1 \dots \mu_n}) + (\hat{X}_{(\partial_\mu \xi^{\mu_1}} \hat{T}_{\nu_1 \dots \nu_n}^{\mu \dots \mu_n}) + \dots + (\hat{X}_{(\partial_\mu \xi^{\mu_n}} \hat{T}_{\nu_1 \dots \nu_n}^{\mu_1 \dots \mu}) \\ & - (\hat{X}_{(\partial_{\nu_1} \xi^\nu} \hat{T}_{\nu \dots \nu_n}^{\mu_1 \dots \mu_n}) - \dots - (\hat{X}_{(\partial_{\nu_n} \xi^\nu} \hat{T}_{\nu_1 \dots \nu}^{\mu_1 \dots \mu_n}). \end{aligned}$$

3.17.1.5 Noncommutative Space-Time Geometry

The deformed algebra of infinitesimal diffeomorphisms and the tensor calculus covariant with respect to it is the fundamental building-block for the definition of a noncommutative geometry on θ -deformed spaces. In this section we sketch the important steps towards a deformed Einstein-Hilbert action [Aschieri *et. al.* (2005)]. A first ingredient is the *covariant derivative* \hat{D}_μ . Algebraically, it can be defined by demanding that acting on a vector-field it produces a tensor-field [Meyer (2005)]

$$\hat{\delta}_\xi \hat{D}_\mu \hat{V}_\nu = -(\hat{X}_\xi \hat{D}_\mu \hat{V}_\nu) - (\hat{X}_{(\partial_\mu \xi^\alpha} \hat{D}_\alpha \hat{V}_\nu) - (\hat{X}_{(\partial_\nu \xi^\alpha} \hat{D}_\mu \hat{V}_\alpha) \quad (3.289)$$

The covariant derivative is given by a *connection* $\hat{\Gamma}_{\mu\nu}{}^\rho$

$$\hat{D}_\mu \hat{V}_\nu = \hat{\partial}_\mu \hat{V}_\nu - \hat{\Gamma}_{\mu\nu}{}^\rho \hat{V}_\rho.$$

From (3.289) it is possible to deduce the transformation property of $\hat{\Gamma}_{\mu\nu}{}^\rho$

$$\hat{\delta}_\xi \hat{\Gamma}_{\mu\nu}{}^\rho = (\hat{X}_\xi \hat{\Gamma}_{\mu\nu}{}^\rho) - (\hat{X}_{(\partial_\mu \xi^\alpha)} \hat{\Gamma}_{\alpha\nu}{}^\rho) - (\hat{X}_{(\partial_\nu \xi^\alpha)} \hat{\Gamma}_{\mu\alpha}{}^\rho) + (\hat{X}_{(\partial_\alpha \xi^\rho)} \hat{\Gamma}_{\mu\nu}{}^\alpha) - (\hat{\partial}_\mu \hat{\partial}_\nu \hat{\xi}^\rho).$$

The *metric* $\hat{G}_{\mu\nu}$ is defined as a symmetric tensor of rank two. It can be obtained for example by a set of vector-fields $\hat{E}_\mu{}^a$, $a = 0, \dots, 3$, where a is to be understood as a mere label. These vector-fields are called *vierbeins*. Then the symmetrized product of those vector-fields is indeed a symmetric tensor of rank two

$$\hat{G}_{\mu\nu} := \frac{1}{2}(\hat{E}_\mu{}^a \hat{E}_\nu{}^b + \hat{E}_\nu{}^b \hat{E}_\mu{}^a) \eta_{ab}.$$

Here η_{ab} stands for the usual flat Minkowski space metric with signature $(-+++)$. Let us assume that we can choose the vierbeins $\hat{E}_\mu{}^a$ such that they reduce in the commutative limit to the usual vierbeins $e_\mu{}^a$. Then also the metric $\hat{G}_{\mu\nu}$ reduces to the usual, undeformed metric $g_{\mu\nu}$.

The inverse metric tensor we denote by upper indices

$$\hat{G}_{\mu\nu} \hat{G}^{\nu\rho} = \delta_\mu^\rho.$$

We use $\hat{G}_{\mu\nu}$ respectively $\hat{G}^{\mu\nu}$ to raise and lower indices.

The curvature and torsion tensors are obtained by taking the commutator of two covariant derivatives¹⁵

$$[\hat{D}_\mu, \hat{D}_\nu] \hat{V}_\rho = \hat{R}_{\mu\nu\rho}{}^\alpha \hat{V}_\alpha + \hat{T}_{\mu\nu}{}^\alpha \hat{D}_\alpha \hat{V}_\rho$$

which leads to the expressions

$$\begin{aligned} \hat{R}_{\mu\nu\rho}{}^\sigma &= \hat{\partial}_\nu \hat{\Gamma}_{\mu\rho}{}^\sigma - \hat{\partial}_\mu \hat{\Gamma}_{\nu\rho}{}^\sigma + \hat{\Gamma}_{\nu\rho}{}^\beta \hat{\Gamma}_{\mu\beta}{}^\sigma - \hat{\Gamma}_{\mu\rho}{}^\beta \hat{\Gamma}_{\nu\beta}{}^\sigma \\ \hat{T}_{\mu\nu}{}^\alpha &= \hat{\Gamma}_{\nu\mu}{}^\alpha - \hat{\Gamma}_{\mu\nu}{}^\alpha. \end{aligned}$$

If we assume the *torsion-free* case, i.e.,

$$\hat{\Gamma}_{\mu\nu}{}^\sigma = \hat{\Gamma}_{\nu\mu}{}^\sigma,$$

we find a unique expression for the metric connection (Christoffel symbols) defined (by means of $\hat{D}_\alpha \hat{G}_{\beta\gamma} = 0$) in terms of the metric and its inverse

$$\hat{\Gamma}_{\alpha\beta}{}^\sigma = \frac{1}{2}(\hat{\partial}_\alpha \hat{G}_{\beta\gamma} + \hat{\partial}_\beta \hat{G}_{\alpha\gamma} - \hat{\partial}_\gamma \hat{G}_{\alpha\beta}) \hat{G}^{\gamma\sigma}.$$

¹⁵The generalization of covariant derivatives acting on tensors is straight forward [Aschieri *et. al.* (2005)].

From the Riemann curvature tensor $\hat{R}_{\mu\nu\rho}{}^\sigma$ we get the Ricci curvature scalar by contracting the indices

$$\hat{R} := \hat{G}^{\mu\nu} \hat{R}_{\nu\mu\rho}{}^\rho.$$

\hat{R} indeed transforms as a scalar which may be checked explicitly by taking the deformed coproduct (3.285) into account.

To get an integral which is invariant with respect to the Hopf algebra of deformed infinitesimal diffeomorphisms we need a measure function \hat{E} . We demand the transformation property

$$\hat{\delta}_\xi \hat{E} = -\hat{X}_\xi \hat{E} - \hat{X}_{(\partial_\mu \xi^\mu)} \hat{E}. \quad (3.290)$$

Then it follows with the deformed coproduct (3.285) that for any scalar field \hat{S}

$$\hat{\delta}_\xi \hat{E} \hat{S} = -\hat{\partial}_\mu (\hat{X}_{\xi^\mu} (\hat{E} \hat{S})).$$

Hence, transforming the product of an arbitrary scalar field with a measure function \hat{E} we get a total derivative which vanishes under the integral. A suitable measure function with the desired transformation property (3.290) is for instance given by the determinant of the vierbein $\hat{E}_\mu{}^a$

$$\hat{E} = \det(\hat{E}_\mu{}^a) := \frac{1}{4!} \varepsilon^{\mu_1 \cdots \mu_4} \varepsilon_{a_1 \cdots a_4} \hat{E}_{\mu_1}{}^{a_1} \hat{E}_{\mu_2}{}^{a_2} \hat{E}_{\mu_3}{}^{a_3} \hat{E}_{\mu_4}{}^{a_4}.$$

That \hat{E} transforms correctly can be shown by using that the product of four $\hat{E}_{\mu_i}{}^{a_i}$ transforms as a tensor of fourth rank and some combinatorics.

Now we have all ingredients to write down the *Einstein–Hilbert action*. Note that having chosen a differential calculus as in (3.271), the integral is uniquely determined up to a normalization factor by requiring¹⁶[Douglas and Nekrasov (2001)]

$$\int \hat{\partial}_\mu \hat{f} = 0$$

for all $\hat{f} \in \hat{\mathcal{A}}$. Then we define the Einstein–Hilbert action on $\hat{\mathcal{A}}$ as

$$\hat{S}_{\text{EH}} := \int \det(\hat{E}_\mu{}^a) \hat{R} + \text{complex conjugate}.$$

¹⁶We consider functions that “vanish at infinity”.

It is by construction invariant with respect to deformed diffeomorphisms meaning that

$$\hat{\delta}_\xi \hat{S}_{\text{EH}} = 0.$$

In this section we have presented the fundamentals of a noncommutative geometry on the algebra $\hat{\mathcal{A}}$ and defined an invariant Einstein–Hilbert action. There is however one important step missing which is subject of the following section: We want to make contact of the noncommutative gravity theory with Einstein’s gravity theory. This we achieve by introducing the \star –product formalism.

3.17.1.6 Star-Products and Expanded Einstein–Hilbert Action

To express the noncommutative fields in terms of their commutative counterparts we first observe that we can map the whole algebraic construction of the previous sections to the algebra of commutative functions via the vector space isomorphism W introduced above. By equipping the algebra of commutative functions with a new product denoted by \star we can render W an algebra isomorphism. We define [Meyer (2005)]

$$f \star g := W^{-1}(W(f)W(g)) = W^{-1}(\hat{f}\hat{g}), \quad (3.291)$$

and get $(\mathcal{A}, \star) \cong \hat{\mathcal{A}}$.

The \star –product corresponding to the symmetric ordering prescription W is then given explicitly by the *Moyal product*¹⁷

$$f \star g = \mu \circ e^{\frac{i}{2}\theta^{\mu\nu}\partial_\mu \otimes \partial_\nu} f \otimes g = fg + \frac{i}{2}\theta^{\mu\nu}(\partial_\mu f)(\partial_\nu g) + \mathcal{O}(\theta^2).$$

It is a deformation of the commutative point–wise product to which it reduces in the limit $\theta \rightarrow 0$.

In virtue of the isomorphism W we can map all noncommutative fields to commutative functions in \mathcal{A}

$$\hat{F} \mapsto W^{-1}(\hat{F}) \equiv F.$$

We then expand the image F in orders of the deformation parameter θ

$$F = F^{(0)} + F^{(1)} + F^{(2)} + \mathcal{O}(\theta^3),$$

¹⁷This is an immediate consequence of (3.275).

where the zeroth order always corresponds to the undeformed quantity. Products of functions in $\hat{\mathcal{A}}$ are simply mapped to \star -products of the corresponding functions in \mathcal{A} . The same can be done for the action of the derivative $\hat{\partial}_\mu$ and consequently for an arbitrary differential operator acting on $\hat{\mathcal{A}}$ [Aschieri *et. al.* (2005)].

The fundamental dynamical field of our gravity theory is the vierbein field \hat{E}_μ^a . All other quantities such as metric, connection and curvature can be expressed in terms of it. Its image with respect to W^{-1} is denoted by E_μ^a . In the first approximation, we study the case $E_\mu^a = e_\mu^a$, where e_μ^a is the usual vierbein field. Then for instance the metric is given up to second order in θ by

$$\begin{aligned} G_{\mu\nu} &= \frac{1}{2}(E_\mu^a \star E_\nu^b + E_\nu^b \star E_\mu^a)\eta_{ab} = \frac{1}{2}(e_\mu^a \star e_\nu^b + e_\nu^b \star e_\mu^a)\eta_{ab} \\ &= g_{\mu\nu} - \frac{1}{8}\theta^{\alpha_1\beta_1}\theta^{\alpha_2\beta_2}(\partial_{\alpha_1}\partial_{\alpha_2}e_\mu^a)(\partial_{\beta_1}\partial_{\beta_2}e_\nu^b)\eta_{ab} + \dots, \end{aligned}$$

where $g_{\mu\nu}$ is the usual, undeformed metric. For the Christoffel symbol one finds up to the second order [Meyer (2005)]:

(i) the 0th order is the undeformed expression

$$\Gamma_{\mu\nu}^{(0)\rho} = \frac{1}{2}[p_\mu g_{\nu\gamma} + p_\nu g_{\mu\gamma} - p_\gamma g_{\mu\nu}]g^{\gamma\rho};$$

(ii) the first order reads

$$\Gamma_{\mu\nu}^{(1)\rho} = \frac{i}{2}\theta^{\alpha\beta}(\partial_\alpha\Gamma_{\mu\nu}^{(0)\sigma})g_{\sigma\tau}(\partial_\beta g^{\tau\rho}); \quad \text{and}$$

(iii) the second order is

$$\begin{aligned} \Gamma_{\mu\nu}^{(2)\rho} &= -\frac{1}{8}\theta^{\alpha_1\beta_1}\theta^{\alpha_2\beta_2}((\partial_{\alpha_1}\partial_{\alpha_2}\Gamma_{\mu\nu\sigma}^{(0)}) (\partial_{\beta_1}\partial_{\beta_2}g^{\sigma\rho}) \\ &\quad - 2(\partial_{\alpha_1}\Gamma_{\mu\nu\sigma}^{(0)})\partial_{\beta_1}((\partial_{\alpha_2}g^{\sigma\tau})(\partial_{\beta_2}g_{\tau\xi})g^{\xi\rho}) \\ &\quad - \Gamma_{\mu\nu\sigma}^{(0)}((\partial_{\alpha_1}\partial_{\alpha_2}g^{\sigma\tau})(\partial_{\beta_1}\partial_{\beta_2}g_{\tau\xi}) + g^{\sigma\tau}(\partial_{\alpha_1}\partial_{\alpha_2}e_\tau^a)(\partial_{\beta_1}\partial_{\beta_2}e_\xi^b)\eta_{ab} \quad (3.292) \\ &\quad - 2\partial_{\alpha_1}((\partial_{\alpha_2}g^{\sigma\tau})(\partial_{\beta_2}g_{\tau\lambda})g^{\lambda\kappa})(\partial_{\beta_1}g_{\kappa\xi})g^{\xi\rho} + \frac{1}{2}(\partial_\mu((\partial_{\alpha_1}\partial_{\alpha_2}e_\nu^a)(\partial_{\beta_1}\partial_{\beta_2}e_\sigma^b)) \\ &\quad + \partial_\nu((\partial_{\alpha_1}\partial_{\alpha_2}e_\sigma^a)(\partial_{\beta_1}\partial_{\beta_2}e_\mu^b)) - \partial_\sigma((\partial_{\alpha_1}\partial_{\alpha_2}e_\mu^a)(\partial_{\beta_1}\partial_{\beta_2}e_\nu^b)))\eta_{ab}g^{\sigma\rho}), \end{aligned}$$

where $\Gamma_{\mu\nu\sigma}^{(0)} = \Gamma_{\mu\nu}^{(0)\rho}g_{\rho\sigma}$.

The expressions for the curvature tensor now read

$$\begin{aligned}
 R_{\mu\nu\rho}^{(1)\sigma} = & -\frac{i}{2}\theta^{\kappa\lambda}((\partial_\kappa R_{\mu\nu\rho}^{(0)\tau})(\partial_\lambda g_{\tau\gamma})g^{\gamma\sigma} - (\partial_\kappa \Gamma_{\nu\rho}^{(0)\beta})(\Gamma_{\mu\beta}^{(0)\tau}(\partial_\lambda g_{\tau\gamma})g^{\gamma\sigma} \\
 & - \Gamma_{\mu\tau}^{(0)\sigma}(\partial_\lambda g_{\beta\gamma})g^{\gamma\tau} + \partial_\mu((\partial_\lambda g_{\beta\gamma})g^{\gamma\sigma}) + (\partial_\lambda \Gamma_{\mu\beta}^{(0)\sigma})) \\
 & + (\partial_\kappa \Gamma_{\mu\rho}^{(0)\beta})(\Gamma_{\nu\beta}^{(0)\tau}(\partial_\lambda g_{\tau\gamma})g^{\gamma\sigma} - \Gamma_{\nu\tau}^{(0)\sigma}(\partial_\lambda g_{\beta\gamma})g^{\gamma\tau} \\
 & + \partial_\nu[(\partial_\lambda g_{\beta\gamma})g^{\gamma\sigma}] + (\partial_\lambda \Gamma_{\nu\beta}^{(0)\sigma})), \tag{3.293}
 \end{aligned}$$

$$\begin{aligned}
 R_{\mu\nu\rho}^{(2)\sigma} = & \partial_\nu \Gamma_{\mu\rho}^{(2)\sigma} + \Gamma_{\nu\rho}^{(2)\gamma} \Gamma_{\mu\gamma}^{(0)\sigma} + \Gamma_{\nu\rho}^{(0)\gamma} \Gamma_{\mu\gamma}^{(2)\sigma} \\
 & + \frac{i}{2}\theta^{\alpha\beta}[(\partial_\alpha \Gamma_{\nu\rho}^{(1)\gamma})(\partial_\beta \Gamma_{\mu\gamma}^{(0)\sigma}) + (\partial_\alpha \Gamma_{\nu\rho}^{(0)\gamma})(\partial_\beta \Gamma_{\mu\gamma}^{(1)\sigma})] \\
 & - \frac{1}{8}\theta^{\alpha_1\beta_1}\theta^{\alpha_2\beta_2}(\partial_{\alpha_1}\partial_{\alpha_2}\Gamma_{\nu\rho}^{(0)\gamma})(\partial_{\beta_1}\partial_{\beta_2}\Gamma_{\mu\gamma}^{(0)\sigma}) - (\mu \leftrightarrow \nu), \tag{3.294}
 \end{aligned}$$

where the second order is given implicitly in terms of the Christoffel symbols.

The *deformed Einstein–Hilbert action* is given by

$$\begin{aligned}
 S_{\text{EH}} &= \frac{1}{2} \int d^4x \det_\star e_\mu^a \star R + \text{c.c.} \\
 &= \frac{1}{2} \int d^4x \det_\star e_\mu^a \star (R + \bar{R}) = \frac{1}{2} \int d^4x \det_\star e_\mu^a (R + \bar{R}) \\
 &= S_{\text{EH}}^{(0)} + \int d^4x (\det e_\mu^a) R^{(2)} + (\det_\star e_\mu^a)^{(2)} R^{(0)}, \tag{3.295}
 \end{aligned}$$

where we used that the integral together with the Moyal product (by partial integration) has the property

$$\int d^4x f \star g = \int d^4x f g = \int d^4x g \star f.$$

In (3.295) $\det_\star e_\mu^a$ is the \star -determinant

$$\det_\star e_\mu^a = \frac{1}{4!} \varepsilon^{\mu_1 \dots \mu_4} \varepsilon_{a_1 \dots a_4} e_{\mu_1}^{a_1} \star e_{\mu_2}^{a_2} \star e_{\mu_3}^{a_3} \star e_{\mu_4}^{a_4} = \det e_\mu^a + (\det_\star e_\mu^a)^{(2)} + \dots,$$

where

$$\begin{aligned}
 (\det_\star)^{(2)} = & -\frac{1}{8} \frac{1}{4!} \theta^{\alpha_1\beta_1} \theta^{\alpha_2\beta_2} \varepsilon^{\mu_1 \dots \mu_4} \varepsilon_{a_1 \dots a_4} [(\partial_{\alpha_1} \partial_{\alpha_2} e_{\mu_1}^{a_1})(\partial_{\beta_1} \partial_{\beta_2} e_{\mu_2}^{a_2}) e_{\mu_3}^{a_3} e_{\mu_4}^{a_4} + \\
 & \partial_{\alpha_1} \partial_{\alpha_2} (e_{\mu_1}^{a_1} e_{\mu_2}^{a_2}) (\partial_{\beta_1} \partial_{\beta_2} e_{\mu_3}^{a_3}) e_{\mu_4}^{a_4} + \partial_{\alpha_1} \partial_{\alpha_2} (e_{\mu_1}^{a_1} e_{\mu_2}^{a_2} e_{\mu_3}^{a_3}) (\partial_{\beta_1} \partial_{\beta_2} e_{\mu_4}^{a_4})].
 \end{aligned}$$

The odd orders of θ vanish in (3.295) but the even orders of θ give nontrivial contributions.

3.17.2 Synthetic Differential Geometry

The sense in which we understand the word ‘synthetic’ in this context is that we place ourselves in the certain category \mathbf{E} of manifold-like objects where everything is smooth. A main assumption about our category \mathbf{E} is that it is cartesian closed, meaning that functional spaces as well as methods of classical functional analysis are available.

Recall that *distributions* are usually thought of as very non-smooth functions, like the Heaviside function, or the Dirac δ -function. These so-called *generalized functions* are commonly presented following the *Sobolev–Schwartz functional analysis*, usually including the integral transforms of Fourier, Laplace, Mellin, Hilbert, Cauchy–Bochner and Poisson. The main application of the theory of generalized functions is the solution of classical *equations of mathematical physics* (see e.g., [Vladimirov (1971); Vladimirov (1986)]).

On the other hand, there is a viewpoint, firstly stressed by Lawvere [Lawvere (1979)], and fully elaborated by A. Kock [Kock (1981); Kock and Reyes (2003)], that distributions are *extensive quantities*, where functions are *intensive quantities*. This viewpoint also makes it quite natural to formulate partial equations of mathematical physics (like classical wave and heat equations) – as ODEs describing the evolution over time of any initial distributions. For example, the main construction in the theory of the wave equation is the construction of the fundamental solution: the description of the evolution of a point δ -distribution over time.

To say that *distributions* are extensive quantities implies that they transform *covariantly* (in a categorical sense). To say that *functions* are intensive quantities implies that they transform *contravariantly*. Distributions are here construed, as linear functionals on the space of (smooth) functions. However, since all functions in the synthetic context are smooth, as well as continuous, there is no distinction between distributions and Radon measures.

In the category \mathbf{E} one can define the vector space $\mathcal{D}'_c(M)$ of distributions of compact support on M , for each manifold-like object $M \in \mathbf{E}$, namely the object of \mathbb{R} -linear maps $\mathbb{R}^M \rightarrow \mathbb{R}$. We shall assume that elementary differential calculus for functions $\mathbb{R} \rightarrow \mathbb{R}$ is available, as in all models of *synthetic differential geometry* (SDG, see [Kock (1981); Moerdijk and Reyes (1991)]). Following Kock [Kock (1981)], we shall also assume some integral calculus, but only in the weakest possible sense, namely we assume that for every $\psi : \mathbb{R} \rightarrow \mathbb{R}$, there is a unique $\Psi : \mathbb{R} \rightarrow \mathbb{R}$ with $\Psi' = \psi$ and with

$\Psi(0) = 0$. Using this result, *intervals* will be construed as *distributions*: for $a, b \in \mathbb{R}$, $[a, b]$ denotes the distribution

$$\psi \mapsto \int_a^b \psi(x) dx = \Psi(b) - \Psi(a).$$

3.17.2.1 Distributions

Let us make the formula for covariant functoriality \mathcal{D}'_c explicit. Let $f : M \rightarrow N$ be a map. Then the corresponding map $f_* = \mathcal{D}'_c(f) : \mathcal{D}'_c(M) \rightarrow \mathcal{D}'_c(N)$ is described by declaring

$$\langle f_*(T), \phi \rangle = \langle T, \phi \circ f \rangle, \quad (3.296)$$

where T is a distribution on M , and ϕ is a function on N . The brackets denote evaluation of distributions on functions. If we similarly denote the value of the contravariant functor $M \mapsto \mathbb{R}^M$ on a map f by f^* , the defining equation for f_* reads

$$\langle f_*(T), \phi \rangle = \langle T, f^*(\phi) \rangle.$$

$\mathcal{D}'_c(M)$ is an \mathbb{R} -linear space, and all maps $f_* : \mathcal{D}'_c(M) \rightarrow \mathcal{D}'_c(N)$ are \mathbb{R} -linear. Also $\mathcal{D}'_c(M)$ is a *Euclidean* vector space V , meaning that the basic differential calculus is available.

For any distribution T of compact support on M , one has its *Total*, which is just the number $\langle T, 1 \rangle \in \mathbb{R}$, where 1 denotes the function on M with constant value 1. Since $f^*(1) = 1$ for any map f , it follows that f_* preserves Totals.

Recall that a distribution T on M may be multiplied by any function $g : M \rightarrow \mathbb{R}$, by the rule

$$\langle g \cdot T, \phi \rangle = \langle T, g \cdot \phi \rangle. \quad (3.297)$$

A basic result from single-variable calculus, ‘integration by substitution’, in a pure ‘distribution’ form reads: Given any function $g : \mathbb{R} \rightarrow \mathbb{R}$, and given $a, b \in \mathbb{R}$, we have

$$g_*(g' \cdot [a, b]) = [g(a), g(b)].$$

Let ψ be a test function, and let Ψ be a primitive of it, $\Psi' = \psi$. Then

$$\langle [g(a), g(b)], \psi \rangle = \Psi(g(b)) - \Psi(g(a)).$$

On the other hand, by the chain rule, $\Psi \circ g$ is a primitive of $g' \cdot (\psi \circ g)$, so

$$\Psi(g(a)) - \Psi(g(b)) = \langle [a, b], g' \cdot (\psi \circ g) \rangle = \langle g' \cdot [a, b], \psi \circ g \rangle = \langle g_* (g' \cdot [a, b]), \psi \rangle.$$

The external product of distributions of compact support is defined as follows. If P is a distribution on M , and Q a distribution on N , we get a distribution $P \times Q$ on $M \times N$, by

$$\langle P \times Q, \psi \rangle = \langle P, [m \mapsto \langle Q, \psi(m, -) \rangle] \rangle.$$

However, if $[a, b]$ and $[c, d]$ are intervals (viewed as distributions on \mathbb{R} , as described above), $[a, b] \times [c, d] = [a, b] \overline{\times} [c, d]$, as distributions on \mathbb{R}^2 , by an application of Fubini's Theorem, (which holds in our context here as a consequence of equality of mixed partial derivatives). Distributions arising in this way on \mathbb{R}^2 , are called *rectangles*. The obvious generalizations to higher dimensions are called *boxes*. We have

$$\langle [a, b] \times [c, d], \psi \rangle = \int_a^b \int_c^d \psi(x, y) \, dy \, dx,$$

in traditional notation. Here, we can define *the boundary* of the box $[a, b] \times [c, d]$ as the obvious distribution on \mathbb{R}^2 ,

$$(p_c^2)_* [a, b] + (p_b^1)_* [c, d] - (p_d^2)_* [a, b] - (p_a^1)_* [c, d],$$

where $p_c^2(x) = (x, c)$, $p_b^1(y) = (b, y)$, etc.

By a *singular* box in a manifold-like object M , we understand the data of a map $\gamma : \mathbb{R}^2 \rightarrow M$ and a box $[a, b] \times [c, d]$ in \mathbb{R}^2 , and similarly for singular intervals and singular rectangles. Such a singular box gives rise to a distribution on M , namely $g_*([a, b] \times [c, d])$.

By *differential operator* on an object M , we here understand just an \mathbb{R} -linear map $D : \mathbb{R}^M \rightarrow \mathbb{R}^M$. If D is such an operator, and T is a distribution on M , we define $D(T)$ by

$$\langle D(T), \psi \rangle = \langle T, D(\psi) \rangle,$$

and in this way, D becomes a linear operator $\mathcal{D}'_c(M) \rightarrow \mathcal{D}'_c(M)$.

In particular, if X is a vector-field on M , one defines the *directional* (i.e., *Lie*) *derivative* $D_X(T)$ of a distribution T on M by the formula

$$\langle D_X(T), \psi \rangle = \langle T, D_X(\psi) \rangle. \quad (3.298)$$

This in particular applies to the vector-field $\partial/\partial x$ on \mathbb{R} , and reads here

$$\langle T', \psi \rangle = \langle T, \psi' \rangle,$$

where ψ' denotes the ordinary derivative of the function ψ .

The following Proposition is an application of the covariant functoriality of the functor \mathcal{D}'_c , which will be used in connection with the wave (and heat) equations in dimension 2. We consider the (orthogonal) projection $p : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ onto the xy -plane; Δ denotes the Laplace operator in the relevant \mathbb{R}^n , so for \mathbb{R}^3 , Δ is $\partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$. For any distribution S (of compact support) on \mathbb{R}^3 ,

$$p_*(\Delta(S)) = \Delta(p_*(S)).$$

This follows from the fact that for any $\psi : \mathbb{R}^2 \rightarrow \mathbb{R}$,

$$\Delta(p^*\psi) = p^*(\Delta(\psi)),$$

namely, $\partial^2\psi/\partial x^2 + \partial^2\psi/\partial y^2$.

3.17.2.2 Synthetic Calculus in Euclidean Spaces

Recall that a *vector space* E (which in the present context means an \mathbb{R} -module) is called *Euclidean* if differential and integral calculus for functions $\mathbb{R} \rightarrow E$ is available (see [Kock (1981); Moerdijk and Reyes (1991)]). The coordinate vector spaces are Euclidean, but so are also the vector spaces \mathbb{R}^M , and $\mathcal{D}'_c(M)$ for any M . To describe for instance the ‘time-derivative’ \dot{f} of a function $f : \mathbb{R} \rightarrow \mathcal{D}'_c(M)$, we put

$$\langle \dot{f}(t), \psi \rangle = \frac{d}{dt} \langle f(t), \psi \rangle.$$

Similarly, from the integration Axiom for \mathbb{R} , one immediately proves that $\mathcal{D}'_c(M)$ satisfies the integration Axiom, in the sense that for any $h : \mathbb{R} \rightarrow \mathcal{D}'_c(M)$, there exists a unique $H : \mathbb{R} \rightarrow \mathcal{D}'_c(M)$ satisfying $H(0) = 0$ and $H'(t) = h(t)$ for all t . In particular, if $h : \mathbb{R} \rightarrow \mathcal{D}'_c(M)$, the ‘integral’ $\int_a^b h(u) du = H(b) - H(a)$ makes sense, and the Fundamental Theorem of Calculus holds.

As a particular case of special importance, we consider a *linear* vector-field on a Euclidean \mathbb{R} -module V . To say that the vector field is *linear* is to say that its principal-part formation $V \rightarrow V$ is a linear map, Γ , say. We have then the following version of a classical result. By a *formal* solution for an ordinary differential equation, we mean a solution defined on the set D_∞ of nilpotent elements in \mathbb{R} (these form a subgroup of $(\mathbb{R}, +)$).

Let a linear vector-field on a Euclidean vector space V be given by the linear map $\Gamma : V \rightarrow V$. Then the unique formal solution of the correspond-

ing differential equation,

$$\dot{F}(t) = \Gamma(F(t)),$$

with initial position v , is the map $D_\infty \times V \rightarrow V$ given by [Kock (1981); Kock and Reyes (2003)]

$$(t, v) \mapsto e^{t \cdot \Gamma}(v), \quad (3.299)$$

where the r.h.s here means the sum of the following ‘series’ (which has only finitely many non-vanishing terms, since t is assumed nilpotent):

$$v + t\Gamma(v) + \frac{t^2}{2!}\Gamma^2(v) + \frac{t^3}{3!}\Gamma^3(v) + \dots$$

where $\Gamma^2(v)$ means $\Gamma(\Gamma(v))$, etc.

There is an analogous result for second order differential equations of the form

$$\ddot{F}(t) = \Gamma(F(t)).$$

The formal solution of this second order differential equation with initial position v and initial velocity w , is given by [Kock (1981); Kock and Reyes (2003)]

$$F(t) = v + t \cdot w + \frac{t^2}{2!}\Gamma(v) + \frac{t^3}{3!}\Gamma(w) + \frac{t^4}{4!}\Gamma^2(v) + \frac{t^5}{5!}\Gamma^2(w) + \dots$$

Also, given $f : \mathbb{R} \rightarrow V$, where V is a Euclidean vector space, and given $g : \mathbb{R} \rightarrow \mathbb{R}$. Then for any $a, b \in \mathbb{R}$,

$$\int_a^b f(g(x)) \cdot g'(x) dx = \int_{g(a)}^{g(b)} f(u) du.$$

Linear maps between Euclidean vector spaces preserve differentiation and integration of functions $\mathbb{R} \rightarrow V$; we shall explicitly need the following particular assertion: Let $F : V \rightarrow W$ be a linear map between Euclidean vector spaces. Then for any $f : \mathbb{R} \rightarrow V$,

$$F\left(\int_a^b f(t) dt\right) = \int_a^b F(f(t)) dt.$$

3.17.2.3 Spheres and Balls as Distributions

For $a, b \in \mathbb{R}$, we let $[a, b]$ denote the distribution $f \mapsto \int_a^b f(x) dx$. Recall that such distributions on the line we call *intervals*; the *length* of an interval $[a, b]$ is defined to be $b - a$. Let $[a_1, b_1]$ and $[a_2, b_2]$ be two such intervals. They are equal as distributions if and only if they have same length, $b_1 - a_1 = b_2 - a_2$ ($= l$, say), and $l \cdot (a_1 - a_2) = 0$ (this then also implies $l \cdot (b_1 - b_2) = 0$).

We shall also consider such ‘balls’ in dimension 2 and 3, where, however, t cannot in general be recovered from the distribution, unless t is strictly positive.

We fix a positive integer n . We shall consider the sphere S_t of radius t , and the ball B_t of radius t , for any $t \in \mathbb{R}$, as distributions on \mathbb{R}^n (of compact support), in the following sense:

$$\begin{aligned} \langle S_t, \psi \rangle &= \int_{S_t} \psi(x) dx = t^{n-1} \int_{S_1} \psi(t \cdot u) du, \\ \langle B_t, \psi \rangle &= \int_{B_t} \psi(x) dx = t^n \int_{B_1} \psi(t \cdot u) du, \end{aligned}$$

where du refers to the surface element of the unit sphere S_1 in the first equation and to the volume element of the unit ball B_1 in the second. The expressions involving \int_{S_t} and \int_{B_t} are to be understood symbolically, unless $t > 0$; if $t > 0$, they make sense literally as integrals over sphere and ball, respectively, of radius t , with dx denoting surface-, resp. volume element. The expression on the right in both equations makes sense for any t , and so the distributions S_t and B_t are defined for all t ; in particular, for nilpotent ones.

It is natural to consider also the following distributions S^t and B^t on \mathbb{R}^n (again of compact support):

$$\begin{aligned} \langle S^t, \psi \rangle &= \int_{S_1} \psi(t \cdot u) du, \\ \langle B^t, \psi \rangle &= \int_{B_1} \psi(t \cdot u) du. \end{aligned}$$

For $t > 0$, they may, modulo factors of the type 4π , be considered as ‘average over S_t ’ and ‘average over B_t ’, respectively, since S^t differs from S_t by a factor t^{n-1} , which is just the surface area of S_t (modulo the factor of type 4π), and similarly for B^t .

Note that $S^1 = S_1$ and $B^1 = B_1$. And also note that the definition of S^t and B^t can be formulated as

$$S^t = H_t(S_1), \quad B^t = H_t(B_1),$$

where $H_t : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the homothetic transformation $u \mapsto t \cdot u$, and where we are using the covariant functoriality of distributions of compact support.

For low dimensions, we shall describe the distributions S_t , B_t , S^t and B^t explicitly:

Dimension 1:

$$\begin{aligned} \langle S_t, \psi \rangle &= \psi(-t) + \psi(t), & \langle B_t, \psi \rangle &= \int_{-t}^t \psi(s) \, ds, \\ \langle S^t, \psi \rangle &= \psi(-t) + \psi(t), & \langle B^t, \psi \rangle &= \int_{-1}^1 \psi(t \cdot s) \, ds. \end{aligned}$$

Dimension 2:

$$\begin{aligned} \langle S_t, \psi \rangle &= \int_0^{2\pi} \psi(t \cos \theta, t \sin \theta) \, t \, d\theta, \\ \langle B_t, \psi \rangle &= \int_0^t \int_0^{2\pi} \psi(s \cos \theta, s \sin \theta) \, s \, d\theta \, ds, \\ \langle S^t, \psi \rangle &= \int_0^{2\pi} \psi(t \cos \theta, t \sin \theta) \, d\theta, \\ \langle B^t, \psi \rangle &= \int_0^1 \int_0^{2\pi} \psi(ts \cos \theta, ts \sin \theta) \, s \, d\theta \, ds. \end{aligned}$$

Dimension 3:

$$\begin{aligned} \langle S_t, \psi \rangle &= \int_0^\pi \int_0^{2\pi} \psi(t \cos \theta \sin \phi, t \sin \theta \sin \phi, t \cos \phi) t^2 \sin \phi \, d\theta \, d\phi, \\ \langle B_t, \psi \rangle &= \int_0^t \int_0^\pi \int_0^{2\pi} \psi(s \cos \theta \sin \phi, s \sin \theta \sin \phi, s \cos \phi) s^2 \sin \phi \, d\theta \, d\phi \, ds, \\ \langle S^t, \psi \rangle &= \int_0^\pi \int_0^{2\pi} \psi(t \cos \theta \sin \phi, t \sin \theta \sin \phi, t \cos \phi) \sin \phi \, d\theta \, d\phi, \\ \langle B^t, \psi \rangle &= \int_0^1 \int_0^\pi \int_0^{2\pi} \psi(ts \cos \theta \sin \phi, ts \sin \theta \sin \phi, ts \cos \phi) s^2 \sin \phi \, d\theta \, d\phi \, ds. \end{aligned}$$

These formulas make sense for all t , whereas set-theoretically S_t and B_t (as point sets) only make good sense for $t > 0$.

3.17.2.4 Stokes Theorem for Unit Sphere

Recall that the main Theorem of vector calculus is the *Stokes Theorem*:

$$\int_{\partial\gamma} \omega = \int_{\gamma} d\omega,$$

for ω an $(n-1)$ -form, γ a suitable n -dimensional figure (with appropriate measure on it) and $\partial\gamma$ its geometric boundary. In the synthetic context, the Theorem holds at least for any singular cubical chain $\gamma : I^n \rightarrow M$ (I^n the n -dimensional coordinate cube), because the Theorem may then be reduced to the fundamental Theorem of calculus, which is the only way integration enters in the elementary synthetic context; measure theory not being available therein (see [Moerdijk and Reyes (1991)] for details). Below, we shall apply the result not only for singular *cubes*, but also for singular *boxes*, like the usual $(\gamma : \mathbb{R}^2 \rightarrow \mathbb{R}^2, [0, 2\pi] \times [0, 1])$, ‘parameterizing the unit disk B by polar coordinates’,

$$\gamma(\theta, r) = (r \cos \theta, r \sin \theta). \quad (3.300)$$

We shall need from vector calculus the *Green–Gauss–Ostrogradsky Divergence Theorem*

$$\text{flux of } \mathbf{F} \text{ over } \partial\gamma = \int_{\gamma} (\text{divergence of } \mathbf{F}),$$

with \mathbf{F} a vector-field, for the geometric ‘figure’ γ = the unit ball in \mathbb{R}^n . For the case of the unit ball in \mathbb{R}^n , the reduction of the Divergence Theorem to Stokes’ Theorem is a matter of the *differential* calculus of vector fields and differential forms. For the convenience of the reader, we recall the case $n = 2$.

Given a vector-field $\mathbf{F}(x, y) = (F(x, y), G(x, y))$ in \mathbb{R}^2 , apply Stokes’ Theorem to the differential form

$$\omega = -G(x, y)dx + F(x, y)dy,$$

for the singular rectangle γ given by (3.300) above.

We also need that the trigonometric functions \cos and \sin should be present. We assume that they are given as part of the data, and that they satisfy $\cos^2 + \sin^2 = 1$, and $\cos' = -\sin$, $\sin' = \cos$. Also as part of the data, we need specified an element $\pi \in \mathbb{R}$ so that $\cos \pi = -1, \cos 0 = 1$.

Then we have

$$\begin{cases} \gamma^*(dx) = \cos \theta dr - r \sin \theta d\theta \\ \gamma^*(dy) = \sin \theta dr + r \cos \theta d\theta \\ \gamma^*(dx \wedge dy) = r (dr \wedge d\theta) \end{cases}.$$

Since

$$d\omega = (\partial G/\partial y + \partial F/\partial x) dx \wedge dy = \operatorname{div}(\mathbf{F}) dx \wedge dy,$$

then

$$\gamma^*(d\omega) = \operatorname{div}(\mathbf{F}) r (dr \wedge d\theta).$$

On the other hand,

$$\gamma^*\omega = (F \sin \theta - G \cos \theta)dr + (F r \cos \theta + G r \sin \theta) d\theta, \quad (3.301)$$

(all F , G , and \mathbf{F} to be evaluated at $(r \cos \theta, r \sin \theta)$). Therefore

$$\int_{\gamma} d\omega = \int_0^{2\pi} \int_0^1 \operatorname{div}(\mathbf{F}) r dr d\theta;$$

this is $\int_{B_1} \operatorname{div}(\mathbf{F}) dA$. On the other hand, by Stokes' Theorem $\int_{\gamma} d\omega = \int_{\partial\gamma} \omega$, which is a curve integral of the 1-form (3.301) around the boundary of the rectangle $[0, 2\pi] \times [0, 1]$. This curve integral is a sum of four terms corresponding to the four sides of the rectangle. Two of these (corresponding to the sides $\theta = 0$ and $\theta = 2\pi$) cancel, and the term corresponding to the side where $r = 0$ vanishes because of the r in $r (dr \wedge d\theta)$, so only the side with $r = 1$ remains, and its contribution is, with the correct orientation,

$$\int_0^{2\pi} (F(\cos \theta, \sin \theta) \cos \theta + G(\cos \theta, \sin \theta) \sin \theta) d\theta = \int_{S_1} \mathbf{F} \cdot \mathbf{n} ds,$$

where \mathbf{n} is the outward unit normal of the unit circle. This expression is the flux of \mathbf{F} over the unit circle, which thus equals the divergence integral calculated above.

3.17.2.5 Time Derivatives of Expanding Spheres

We now combine vector calculus with the calculus of the basic ball- and sphere-distributions, to get the following result [Kock and Reyes (2003)]:

In \mathbb{R}^n (for any n), we have, for any t ,

$$\frac{d}{dt} S^t = t \cdot \Delta(B^t),$$

where Δ denotes the *Laplacian operator* (see [Kock (2001)] for ‘synthetic’ analysis of the Laplacian).

We collect some further information about t -derivatives of some of the t -parameterized distributions considered:

$$\frac{d}{dt}(B_t) = S_t. \quad (3.302)$$

In dimension 1, we have

$$\frac{d}{dt}(S_t) = \Delta(B_t), \quad (3.303)$$

for

$$\frac{d}{dt} \langle S_t, \psi \rangle = \frac{d}{dt} \langle \psi(t) + \psi(-t) \rangle = \psi'(t) - \psi'(-t),$$

whereas

$$\langle \Delta B_t, \psi \rangle = \langle B_t, \psi'' \rangle = \int_{-t}^t \psi''(t) dt,$$

and the result follows from the *Fundamental Theorem of calculus*.

The equation (3.303) implies the following equation if $n = 1$, while in [Kock and Reyes (2003)] it was proved that it also holds if $n \geq 2$:

$$t \cdot \frac{d}{dt}(S_t) = (n-1)S_t + t \cdot \Delta(B_t). \quad (3.304)$$

3.17.2.6 The Wave Equation

Let Δ denote the Laplacian operator $\sum \partial^2 / \partial x_i^2$ on \mathbb{R}^n (see [Kock (2001)]). We shall consider the wave equation (WE) in \mathbb{R}^n , (for $n = 1, 2, 3$),

$$\frac{d^2}{dt^2} Q(t) = \Delta Q(t) \quad (3.305)$$

as a second order ordinary differential equation on the Euclidean vector space $\mathcal{D}'_c(\mathbb{R}^n)$ of distributions of compact support; in other words, we are looking for functions

$$Q : \mathbb{R} \rightarrow \mathcal{D}'_c(\mathbb{R}^n),$$

so that for all $t \in \mathbb{R}$, $\ddot{Q}(t) = \Delta(Q(t))$, viewing Δ as a map $\mathcal{D}'_c(\mathbb{R}^n) \rightarrow \mathcal{D}'_c(\mathbb{R}^n)$ [Kock and Reyes (2003)].

Consider a function $f : \mathbb{R} \rightarrow V$, where V is a Euclidean vector space (we are interested in $V = \mathcal{D}'_c(\mathbb{R}^n)$). Then we call the pair of vectors in V

consisting of $f(0)$ and $\dot{f}(0)$ the *initial state* of f . We can now, for each of the cases $n = 1$, $n = 3$, and $n = 2$ describe fundamental solutions to the wave equations. By *fundamental solutions*, we mean solutions whose initial state is either a constant times $(\delta(\underline{0}), 0)$, or a constant times $(0, \delta(\underline{0}))$.

In dimension 1 : The function $\mathbb{R} \rightarrow \mathcal{D}'_c(\mathbb{R})$ given by

$$t \mapsto S^t (= S_t)$$

is a solution of the WE; its initial state is $2(\delta(\underline{0}), 0)$.

The function $\mathbb{R} \rightarrow \mathcal{D}'_c(\mathbb{R})$ given by

$$t \mapsto B_t$$

is a solution of the WE with initial state $2(0, \delta(\underline{0}))$.

In dimension 3: The function $\mathbb{R} \rightarrow \mathcal{D}'_c(\mathbb{R}^3)$ given by

$$t \mapsto S^t + t^2 \Delta(B^t)$$

is a solution of the WE with initial state $4\pi\delta(\underline{0}), 0)$. The function $\mathbb{R} \rightarrow \mathcal{D}'_c(\mathbb{R}^3)$ given by

$$t \mapsto t \cdot S^t$$

is a solution of the WE with initial state $4\pi(0, \delta(\underline{0}))$.

In dimension 2: The function $\mathbb{R} \rightarrow \mathcal{D}'_c(\mathbb{R}^2)$ given by

$$t \mapsto p_*(S^t + t^2 \Delta(B^t))$$

is a fundamental solution of the WE in dimension 2; its initial state is $4\pi(\delta(\underline{0}), 0)$. The function $\mathbb{R} \rightarrow \mathcal{D}'_c(\mathbb{R}^2)$ given by

$$t \mapsto p_*(t \cdot S^t)$$

is a fundamental solution of the WE in dimension 2; its initial state is $4\pi(0, \delta(\underline{0}))$.

This page intentionally left blank

Chapter 4

Applied Bundle Geometry

4.1 Intuition Behind a Fibre Bundle

Recall that tangent and cotangent bundles, TM and T^*M , are special cases of a more general geometrical object called *fibre bundle*, where the word *fiber* V of a map $\pi : Y \rightarrow X$ denotes the *preimage* $\pi^{-1}(x)$ of an element $x \in X$. It is a space which *locally* looks like a product of two spaces (similarly as a manifold locally looks like Euclidean space), but may possess a different *global* structure. To get a visual intuition behind this fundamental geometrical concept, we can say that a fibre bundle Y is a *homeomorphic generalization* of a *product space* $X \times V$ (see Figure 4.1), where X and V are called the *base* and the *fibre*, respectively. $\pi : Y \rightarrow X$ is called the *projection*, $Y_x = \pi^{-1}(x)$ denotes a fibre over a point x of the base X , while the map $f = \pi^{-1} : X \rightarrow Y$ defines the *cross-section*, producing the *graph* $(x, f(x))$ in the bundle Y (e.g., in case of a tangent bundle, $f = \dot{x}$ represents a velocity vector-field) (see [Steenrod (1951)]).

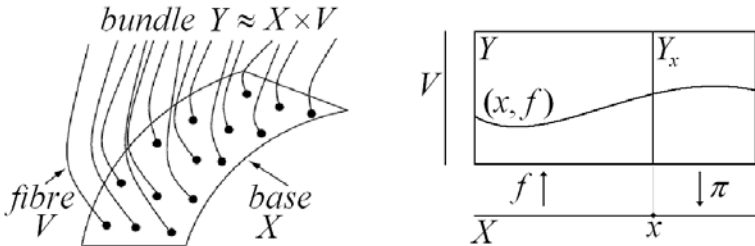


Fig. 4.1 A sketch of a fibre bundle $Y \approx X \times V$ as a generalization of a product space $X \times V$; left – main components; right – a few details (see text for explanation).

The main reason why we need to study fibre bundles is that *all dy-*

namical objects (including vectors, tensors, differential forms and gauge potentials) are their *cross-sections*, representing *generalizations of graphs of continuous functions*.

4.2 Definition of a Fibre Bundle

Let M denote an n -manifold with an atlas Ψ_M consisting of local coordinates $x^\alpha \in M$, ($\alpha = 1, \dots, \dim M$), given by

$$\Psi_M = \{U_\xi, \phi_\xi\}, \quad \phi_\xi(x) = x^\alpha e_\alpha, \quad (\text{for all } x \in U_\xi \subset M),$$

where $\{e_\alpha\}$ is a fixed basis of \mathbb{R}^m . Its tangent and cotangent bundles, TM and T^*M , respectively, admit atlases of induced coordinates $(x^\alpha, \dot{x}^\alpha)$ and $(x^\alpha, \dot{x}_\alpha)$, relative to the *holonomic fibre bases* $\{\partial_\alpha\}$ and $\{dx^\alpha\}$, respectively. For all elements (i.e., points) $p \in TM$ and $p^* \in T^*M$, we have (see [Sardanashvily (1993); Sardanashvily (1995); Giachetta *et. al.* (1997); Mangiarotti and Sardanashvily (2000a); Sardanashvily (2002a)])

$$p = \dot{x}^\alpha \partial_\alpha, \quad p^* = \dot{x}_\alpha dx^\alpha, \quad \partial_\alpha \rfloor dx^\alpha = \delta_\alpha^\alpha, \quad (\alpha = 1, \dots, \dim M).$$

Also, we will use the notation

$$\omega = dx^1 \wedge \dots \wedge dx^n, \quad \omega_\alpha = \partial_\alpha \rfloor \omega, \quad \omega_{\mu\alpha} = \partial_\mu \rfloor \partial_\alpha \rfloor \omega. \quad (4.1)$$

If $f : M \rightarrow M'$ is a smooth manifold map, we define the induced *tangent map* Tf over f , given by

$$Tf : TM \rightarrow TM', \quad \dot{x}'^\alpha \circ Tf = \frac{\partial f^\alpha}{\partial x^\alpha} \dot{x}^\alpha. \quad (4.2)$$

Given a *manifold product* $M \times N$, π_1 and π_2 denote the *natural projections* (i.e., canonical surjections),

$$\pi_1 : M \times N \rightarrow M, \quad \pi_2 : M \times N \rightarrow N.$$

Now, as a homeomorphic generalization of a product space, a fibre bundle can be viewed either as a topological or a geometrical (i.e., coordinate) construction. As a topological construction, a fibre bundle is a class of more general *fibrations*. To have a glimpse of this construction, let $I = [0, 1]$. A map $\pi : Y \rightarrow X$ is said to have the *homotopy lifting property* (HLP, see [Switzer (1975)]) with respect to a topological space Z if for every map $f : Z \rightarrow Y$ and homotopy $H : Z \times I \rightarrow X$ of $\pi \circ f$ there is a homotopy $V : Z \times I \rightarrow Y$ with $f = V_0$ and $\pi \circ V = H$. V is said to be a *lifting* of H . π

is called a *fibration* if it has the HLP for all spaces Z and a *weak fibration* if it has the HLP for all disks D^n , ($n \geq 0$). If $x \in X$ is the base point, then $V = \pi^{-1}(x)$ is called the *fibre* of the fibration π . The projection onto the first factor, $\pi_1 : X \times V \rightarrow X$, is clearly a fibration and is called the *trivial fibration* over X with the fibre V .

However, for the sake of applying differential and integral dynamics on fibre bundles, we will rather use Steenrod's *coordinate bundle definition* (see [Steenrod (1951)]), which defines fibre bundle Y as a sextuple $(Y, X, \pi, V, G, \Psi_Y)$, with:

- (1) a space Y called the *total space*, *bundle space* (or simply *bundle*),
- (2) a space X called the *base space*,
- (3) a surjection $\pi : Y \rightarrow X$ called the *projection*,
- (4) a space $V \subset Y$ called the *fibre*,
- (5) an effective topological (or Lie) transformation group G of V called the *group of the bundle*, and
- (6) a *bundle atlas* Ψ_Y .

Some standard examples of fibre bundles include any *Cartesian product* $X \times V \rightarrow X$ (which is a bundle over X with fibre V), the *Möbius strip* (which is a nontrivial fibre bundle over the circle S^1 with fibre given by the unit interval $I = [0, 1]$; the corresponding trivial bundle is a cylinder), the *Klein bottle* (which can be viewed as a 'twisted' circle bundle over another circle; thus, the corresponding trivial bundle is a torus, $S^1 \times S^1$), a *3-sphere* S^3 (which is a bundle over S^2 with fibre S^1 ; more generally, a *sphere bundle* is a fiber bundle whose fiber is an n -sphere), while a *covering space* is a fiber bundle whose fiber is a discrete space.

Main properties of graphs of functions $f : X \rightarrow V$ carry over to fibre bundles. A graph of such a function, $(x, f(x))$, sits in the product space $X \times V$, or in its homeomorphic generalization bundle. A graph is always 1-1 and projects *onto* the base X .

A special class of fibre bundle is the *vector bundle*, in which the fibre is a vector space. Special cases of fibre bundles that we will use in dynamics of complex systems are: vector, affine, and principal bundles.

A fibre bundle also comes with a group G action on its fibre V , so it can also be called a G -bundle. This group action represents the different ways the fibre V can be viewed as equivalent (e.g., the group G might be the group of homeomorphisms (topological group) or diffeomorphisms (Lie group) of the fibre V ; or, the group G on a vector bundle is the group of invertible linear maps, which reflects the equivalent descriptions of a vector

space using different vector-space bases). A *principal bundle* is G -bundle where the fiber can be identified with the group G itself and where there is a right action of G on the bundle space which is fiber preserving.

Fibre bundles are not always used to generalize functions. Sometimes they are convenient descriptions of interesting manifolds. A common example is a torus bundle on the circle.

More specifically, a fibre bundle, or *fibre bundle* Y over an n D base X is defined as a manifold surjection

$$\pi : Y \rightarrow X, \quad (4.3)$$

where Y admits an atlas Ψ_Y of fibre coordinates

$$(x^\alpha, y^i), \quad x^\alpha \rightarrow x'^\alpha(x^\mu), \quad y^i \rightarrow y'^i(x^\mu, y^j), \quad (4.4)$$

compatible with the fibration (4.3), i.e., such that x^α are coordinates on the base X ,

$$\pi : Y \ni (x^\alpha, y^i) \mapsto x^\alpha \in X.$$

This condition is equivalent to π being a submersion, which means that its tangent map $T\pi : TY \rightarrow TX$ is a surjection. This also implies that π is an *open map*.

A fibre bundle $Y \rightarrow X$ is said to be *trivial* if it is equivalent to the Cartesian product of manifolds, $Y \cong X \times V$, i.e., defined as $\pi_1 : X \times V \rightarrow X$.¹ A fibre bundle over a *contractible base* is always trivial [Steenrod (1951)].

A fibre bundle $Y \rightarrow X$ is said to be *locally trivial* if there exists a fibred coordinate atlas Ψ_Y over an open covering $\{\pi^{-1}(U_\xi)\} \in Y$ of the bundle space Y where $\{U_\xi\} \in X$ is an open covering of the base space X . In other words, all points of the same fibre $Y_x = \pi^{-1}(x)$ of a bundle Y can be covered by the same fibred coordinate chart $\psi_\xi \in \Psi_Y$, so that we have the standard fibre-manifold V for all local bundle splittings

$$\psi_\xi : \pi^{-1}(U_\xi) \rightarrow U_\xi \times V.$$

For the purpose of our general dynamics, the most important fibre bundles are those which are at the same time smooth manifolds. A fibre bundle $Y \rightarrow X$ is said to be *smooth* (C^∞) if there exist a typical fibre-manifold

¹A trivial fibre bundle admits different trivializations $Y \cong X \times V$ that differ from each other in surjections $Y \rightarrow V$.

V and an open covering $\{U_\xi\}$ of X such that Y is locally diffeomorphic to the splittings

$$\psi_\xi : \pi^{-1}(U_\xi) \rightarrow U_\xi \times V, \quad (4.5)$$

glued together by means of smooth *transition functions*

$$\rho_{\xi\zeta} = \psi_\xi \circ \psi_\zeta^{-1} : U_\xi \cap U_\zeta \times V \rightarrow U_\xi \cap U_\zeta \times V \quad (4.6)$$

on overlaps $U_\xi \cap U_\zeta$. It follows that fibres $Y_x = \pi^{-1}(x)$, (for all $x \in X$), of a fibre bundle are its closed imbedded submanifolds. Transition functions $\rho_{\xi\zeta}$ fulfil the *cocycle condition*

$$\rho_{\xi\zeta} \circ \rho_{\zeta\iota} = \rho_{\xi\iota} \quad (4.7)$$

on all overlaps $U_\xi \cap U_\zeta \cap U_\iota$. Trivialization charts (U_ξ, ψ_ξ) together with transition functions $\rho_{\xi\zeta}$ (4.6) constitute a bundle atlas

$$\Psi_Y = \{(U_\xi, \psi_\xi), \rho_{\xi\zeta}\} \quad (4.8)$$

of a fibre bundle $Y \rightarrow X$. Two bundle atlases are said to be *equivalent* if their union is also a bundle atlas, i.e., there exist unique transition functions between trivialization charts of different atlases. A fibre bundle $Y \rightarrow X$ is uniquely defined by a bundle atlas, and all its atlases are equivalent. Every smooth fibre bundle $Y \rightarrow X$ admits a bundle atlas Ψ_Y over a finite covering $\{U_\xi\}$ of X .

If $Y \rightarrow X$ is a fibre bundle, the fibre coordinates $(x^\alpha, y^i) \in Y$ are assumed to be bundle coordinates associated with a bundle atlas Ψ_Y , that is,

$$y^i(y) = (v^i \circ \pi_2 \circ \psi_\xi)(y), \quad (\pi(y) \in U_\xi \subset X), \quad (4.9)$$

where $v^i \in V \subset Y$ are coordinates of the standard fibre V of Y .

Maps of fibre bundles (or, *bundle maps*), by definition, *preserve their fibrations*, i.e., send a fibre to a fibre. Namely, a *bundle map* of a fibre bundle $\pi : Y \rightarrow X$ to a fibre bundle $\pi' : Y' \rightarrow X'$ is defined as a pair (Φ, f) of manifold maps such that the following diagram commutes

$$\begin{array}{ccc} Y & \xrightarrow{\Phi} & Y' \\ \pi \downarrow & & \downarrow \pi' \\ X & \xrightarrow{f} & X' \end{array}$$

i.e., Φ is a fibrewise map over f which sends a fibre Y_x , (for all $x \in X$), to a fibre $Y'_{f(x)}$, (for all $f(x) \in X'$). A *bundle diffeomorphism* is called an *automorphism* if it is an *isomorphism* to itself. In field theory, any automorphism of a fibre bundle is treated as a *gauge transformation*.

Given a bundle $Y \rightarrow X$, every map $f : X' \rightarrow X$ induces a bundle $Y' = f^*Y$ over X' which is called the *pull-back* of the bundle Y by f , such that the following diagram commutes

$$\begin{array}{ccc} Y & \xleftarrow{f^*} & Y' \\ \pi \downarrow & & \downarrow \pi' \\ X & \xleftarrow{f} & X' \end{array}$$

In particular, the product $Y \times Y'$ over X of bundles $\pi : Y \rightarrow X$ and $\pi' : Y' \rightarrow X$ is the pull-back

$$Y \times Y' = \pi^*Y' = \pi'^*Y.$$

Classical fields are described by *sections* of fibre bundles. A (*global*) *section* of a fibre bundle $Y \rightarrow X$ is defined as a π -inverse manifold injection $s : X \rightarrow Y$, $s(x) \mapsto Y_x$, such that $\pi \circ s = \text{Id}_X$. That is, a section s sends any point $x \in X$ into the fibre $Y_x \subset Y$ over this point. A section s is an imbedding, i.e., $s(X) \subset Y$ is both a submanifold and a topological subspace of Y . It is also a closed map, which sends closed subsets of X onto closed subsets of Y . Similarly, a section of a fibre bundle $Y \rightarrow X$ over a submanifold of X is defined. Given a bundle atlas Ψ_Y and associated bundle coordinates (x^α, y^i) , a section s of a fibre bundle $Y \rightarrow X$ is represented by collections of local functions $\{s^i = y^i \circ \psi_\xi \circ s\}$ on trivialization sets $U_\xi \subset X$.

A fibre bundle $Y \rightarrow X$ whose typical fibre is diffeomorphic to an Euclidean space \mathbb{R}^m has a global section. More generally, its section over a closed imbedded submanifold (e.g., a point) of X is extended to a global section [Steenrod (1951)].

In contrast, by a *local section* is usually meant a section over an open subset of the base X . A fibre bundle admits a local section around each point of its base, but need not have a global section.

For any $n \geq 1$ the *normal bundle* NS^n of the n -sphere S^n is the fibre bundle $(S^n, p', E', \mathbb{R}^1)$, where $E' = \{(x, y) \in \mathbb{R}^{n+1} \times \mathbb{R}^{n+1} : \|x\| = 1, y = \lambda x, \lambda \in \mathbb{R}^1\}$ and $p' : E' \rightarrow S^n$ is defined by $p'(x, y) = x$ [Switzer (1975)].

4.3 Vector and Affine Bundles

The most important fibre bundles are vector and affine bundles, which give a standard framework in both classical and quantum dynamics and field theory (e.g., matter fields are sections of vector bundles, while gauge potentials are sections of affine bundles).

Recall that both the tangent bundle (TM, π_M, M) and the cotangent bundle (T^*M, π_M^*, M) are examples of a more general notion of *vector bundle* (E, π, M) of a manifold M , which consists of manifolds E (the total space) and M (the base), as well as a smooth map $\pi : E \rightarrow M$ (the projection) together with an equivalence class of vector bundle atlases (see [Kolar *et al.* (1993)]). A vector bundle atlas $(U_\alpha, \phi_\alpha)_{\alpha \in A}$ for (E, π, M) is a set of pairwise compatible vector bundle charts (U_α, ϕ_α) such that $(U_\alpha)_{\alpha \in A}$ is an open cover of M . Two vector bundle atlases are called equivalent, if their union is again a vector bundle atlas.

On each fibre $E_m = \pi^{-1}(m)$ corresponding to the point $m \in M$ there is a unique structure of a real vector space, induced from any vector bundle chart (U_α, ϕ_α) with $m \in U_\alpha$. A section u of (E, π, M) is a smooth map $u : M \rightarrow E$ with $\pi \circ u = \text{Id}_M$.

Let (E, π_M, M) and (F, π_N, N) be vector bundles. A *vector bundle homomorphism* $\Phi : E \rightarrow F$ is a fibre respecting, fibre linear smooth map induced by the smooth map $\varphi : M \rightarrow N$ between the base manifolds M and N , i.e., the following diagram commutes:

$$\begin{array}{ccc} E & \xrightarrow{\Phi} & F \\ \pi_M \downarrow & & \downarrow \pi_N \\ M & \xrightarrow{\varphi} & N \end{array}$$

We say that Φ *covers* φ . If Φ is invertible, it is called a *vector bundle isomorphism*.

All smooth vector bundles together with their homomorphisms form a category \mathcal{VB} .

If (E, π, M) is a vector bundle which admits a vector bundle atlas $(U_\alpha, \phi_\alpha)_{\alpha \in A}$ with the given open cover, then, we have $\phi_\alpha \circ \phi_\beta^{-1}(m, v) = (m, \phi_{\alpha\beta}(m)v)$ for C^k -transition functions $\phi_{\alpha\beta} : U_{\alpha\beta} = U_\alpha \cap U_\beta \rightarrow GL(V)$ (where we have fixed a standard fibre V). This family of transition maps

satisfies the *cocycle condition*

$$\begin{cases} \phi_{\alpha\beta}(m) \cdot \phi_{\beta\gamma}(m) = \phi_{\alpha\gamma}(m) & \text{for each } m \in U_{\alpha\beta\gamma} = U_\alpha \cap U_\beta \cap U_\gamma, \\ \phi_{\alpha\alpha}(m) = e & \text{for all } m \in U_\alpha. \end{cases}$$

The family $(\phi_{\alpha\beta})$ is called the *cocycle* of transition maps for the vector bundle atlas (U_α, ϕ_α) .

Now, let us suppose that the same vector bundle (E, π, M) is described by an equivalent vector bundle atlas $(U_\alpha, \psi_\alpha)_{\alpha \in A}$ with the same open cover (U_α) . Then the vector bundle charts (U_α, ϕ_α) and (U_α, ψ_α) are compatible for each α , so $\psi_\alpha \circ \phi_\beta^{-1}(m, v) = (m, \tau_\alpha(m)v)$ for some $\tau_\alpha : U_\alpha \rightarrow GL(V)$. We get

$$\tau_\alpha(m) \phi_{\alpha\beta}(m) = \phi_{\alpha\beta}(m) \tau_\beta(m) \quad \text{for all } m \in U_{\alpha\beta},$$

and we say that the two cocycles $(\phi_{\alpha\beta})$ and $(\psi_{\alpha\beta})$ of transition maps over the cover (U_α) are *cohomologous*. If $GL(V)$ is an Abelian group, i.e., if the standard fibre V is of real or complex dimension 1, then the cohomology classes of cocycles $(\phi_{\alpha\beta})$ over the open cover (U_α) form a usual cohomology group $H^1(M, GL(V))$ with coefficients in the sheaf $GL(V)$ [Kolar *et al.* (1993)].

Let (E, π, M) be a vector bundle and let $\varphi : N \rightarrow M$ be a smooth map between the base manifolds N and M . Then there exists the *pull-back vector bundle* $(\varphi^*E, \varphi^*\pi, \varphi^*N)$ with the same typical fibre and a vector bundle homomorphism, given by the commutative diagram [Kolar *et al.* (1993)]:

$$\begin{array}{ccc} \varphi^*E & \xrightarrow{\pi^*\varphi} & E \\ \varphi^*\pi \downarrow & & \downarrow \pi \\ N & \xrightarrow{\varphi} & M \end{array}$$

The vector bundle $(\varphi^*E, \varphi^*\pi, \varphi^*N)$ is constructed as follows. Let $E = VB(\phi_{\alpha\beta})$ denote that E is described by a cocycle $(\phi_{\alpha\beta})$ of transition maps over an open cover (U_α) of M . Then $(\phi_{\alpha\beta} \circ \varphi)$ is a cocycle of transition maps over the open cover $(\varphi^{-1}(U_\alpha))$ of N and the bundle is given by $\varphi^*E = VB(\phi_{\alpha\beta} \circ \varphi)$.

In other words, a vector bundle is a fibre bundle which admits an atlas of linear bundle coordinates. Typical fibres of a smooth vector bundle $\pi : Y \rightarrow X$ are *vector spaces* of some finite dimension (called the fibre

dimension, $\text{fdim } Y$ of Y), and Y admits a bundle atlas Ψ_Y (4.8) where trivialization maps $\psi_\xi(x)$ and transition functions $\rho_{\xi\zeta}(x)$ are linear isomorphisms of vector spaces. The corresponding bundle coordinates (y^i) obey a *linear* coordinate transformation law

$$y'^i = \rho_j^i(x) y^j.$$

We have the decomposition $y = y^i e_i(\pi(y))$, where

$$\{e_i(x)\} = \psi_\xi^{-1}(x)\{v_i\}, \quad x = \pi(y) \in U_\xi,$$

are fibre bases (or frames) for fibres Y_x of Y and $\{v_i\}$ is a fixed basis for the typical fibre V of Y .

There are several standard constructions of new vector bundles from old ones:

- Given two vector bundles Y and Y' over the same base X , their *Whitney sum* $Y \oplus Y'$ is a vector bundle over X whose fibres are the direct sums of those of the vector bundles Y and Y' .
- Given two vector bundles Y and Y' over the same base X , their tensor product $Y \otimes Y'$ is a vector bundle over X whose fibres are the tensor products of those of the vector bundles Y and Y' . In a similar way the exterior product $Y \wedge Y$ of vector bundles is defined, so that the *exterior bundle* of Y is defined as

$$\wedge Y = X \times \mathbb{R} \oplus Y \oplus \wedge^2 Y \oplus \cdots \oplus \wedge^m Y, \quad (m = \text{fdim } Y).$$

- Let $Y \rightarrow X$ be a vector bundle. By $Y^* \rightarrow X$ is denoted the *dual vector bundle* whose fibres are the duals of those of Y . The *interior product* (or *contraction*) of Y and Y^* is defined as a bundle map

$$\lrcorner : Y \otimes Y^* \rightarrow X \times \mathbb{R}.$$

Given a linear bundle map $\Phi : Y' \rightarrow Y$ of vector bundles over X , its kernel $\text{Ker } \Phi$ is defined as the inverse image $\Phi^{-1}(\widehat{0}(X))$ of the canonical zero section $\widehat{0}(X)$ of Y . If Φ is of constant rank, its kernel $\text{Ker } \Phi$ and its image $\text{Im } \Phi$ are subbundles of the vector bundles Y' and Y , respectively. For example, monomorphisms and epimorphisms of vector bundles fulfil this condition. If Y' is a subbundle of the vector bundle $Y \rightarrow X$, the factor bundle Y/Y' over X is defined as a vector bundle whose fibres are the quotients Y_x/Y'_x , $x \in X$.

Consider the *short exact sequence* of vector bundles over X ,

$$0 \rightarrow Y' \xrightarrow{i} Y \xrightarrow{j} Y'' \rightarrow 0, \quad (4.10)$$

which means that i is a bundle monomorphism, j is a bundle epimorphism, and $\text{Ker } j = \text{Im } i$. Then Y'' is the factor bundle Y/Y' . One says that the short exact sequence (4.10) admits a *splitting* if there exists a bundle monomorphism $s : Y'' \rightarrow Y$ such that $j \circ s = \text{Id}_{Y''}$, i.e.,

$$Y = i(Y') \oplus s(Y'') \cong Y' \oplus Y''.$$

Vector bundles of rank 1 are called *line bundles*.

The only two vector bundles with base space B a circle and 1D fibre F are the Möbius band and the annulus, but the classification of all the different vector bundles over a given base space with fibre of a given dimension is quite difficult in general. For example, when the base space is a high-dimensional sphere and the dimension of the fibre is at least three, then the classification is of the same order of difficulty as the fundamental but still largely unsolved problem of computing the homotopy groups of spheres [Hatcher (2002)].

Now, there is a natural direct sum operation for vector bundles over a fixed base space X , which in each fibre reduces just to direct sum of vector spaces. Using this, one can get a weaker notion of isomorphism of vector bundles by defining two vector bundles over the same base space X to be stably isomorphic if they become isomorphic after direct sum with product vector bundles $X \times \mathbb{R}^n$ for some n , perhaps different n 's for the two given vector bundles. Then it turns out that the set of stable isomorphism classes of vector bundles over X forms an Abelian group under the direct sum operation, at least if X is compact Hausdorff. The traditional notation for this group is $\widetilde{KO}(X)$. It is the basis for *K-theory* (see below). In the case of spheres the groups $\widetilde{KO}(S^n)$ have the quite unexpected property of being periodic in n . This is called *Bott Periodicity*, and the values of $\widetilde{KO}(S^n)$ are given by the following table [Hatcher (2002)]:

$n \bmod 8$	1	2	3	4	5	6	7	8
$\widetilde{KO}(S^n)$	\mathbb{Z}_2	\mathbb{Z}_2	0	\mathbb{Z}	0	0	0	\mathbb{Z}

For example, $\widetilde{KO}(S^1)$ is \mathbb{Z}_2 , a cyclic group of order two, and a generator for this group is the Möbius bundle. This has order two since the direct sum of two copies of the Möbius bundle is the product $S^1 \times \mathbb{R}^1$, as one can see by embedding two Möbius bands in a solid torus so that they intersect

orthogonally along the common core circle of both bands, which is also the core circle of the solid torus.

The complex version of $\widetilde{KO}(X)$, called $\widetilde{K}(X)$, is constructed in the same way as $\widetilde{KO}(X)$ but using vector bundles whose fibers are vector spaces over \mathbb{C} rather than \mathbb{R} . The complex form of Bott Periodicity asserts that $\widetilde{K}(S^n)$ is \mathbb{Z} for n even and 0 for n odd, so the period is two rather than eight.

The groups $\widetilde{K}(X)$ and $\widetilde{KO}(X)$ for varying X share certain formal properties with the cohomology groups studied in classical algebraic topology [Hatcher (2002)]. Using a more general form of Bott periodicity, it is in fact possible to extend the groups $\widetilde{K}(X)$ and $\widetilde{KO}(X)$ to a full cohomology theory, families of Abelian groups $\widetilde{K}^n(X)$ and $\widetilde{KO}^n(X)$ for $n \in \mathbb{Z}$ that are periodic in n of period two and eight, respectively. However, there is more algebraic structure here than just the additive group structure. Namely, tensor products of vector spaces give rise to tensor products of vector bundles, which in turn give product operations in both real and complex K -theory similar to cup product in ordinary cohomology. Furthermore, exterior powers of vector spaces give natural operations within K -theory (for more development, see next section, below).

4.3.1 The Second Vector Bundle of the Manifold M

Let (E, π, M) be a vector bundle over the biodynamical manifold M with fibre addition $+_E : E \times_M E \rightarrow E$ and fibre scalar multiplication $m_t^E : E \rightarrow E$. Then (TE, π_E, E) , the tangent bundle of the manifold E , is itself a vector bundle, with fibre addition denoted by $+_{TE}$ and scalar multiplication denoted by m_t^{TE} . The *second vector bundle* structure on $(TE, T\pi, TM)$, is the ‘derivative’ of the original one on (E, π, M) . In particular, the space $\{\Xi \in TE : T\pi_*\Xi = 0 \in TM\} = (Tp)^{-1}(0)$ is denoted by VE and is called the *vertical bundle* over E . Its main characteristics are *vertical lift* and *vertical projection* (see [Kolar *et al.* (1993)] for details).

All of this is valid for the *second tangent bundle* $T^2M = TTM$ of a manifold, but here we have one more natural structure at our disposal. The *canonical flip* or *involution* $\kappa_M : T^2M \rightarrow T^2M$ is defined locally by

$$(T^2\phi \circ \kappa_M \circ T^2\phi^{-1})(x, \xi; \eta, \zeta) = (x, \eta; \xi, \zeta).$$

where (U, ϕ) is a local chart on M (this definition is invariant under changes of charts). The flip κ_M has the following properties (see [Kolar *et al.* (1993)]):

- (1) $\kappa_M \circ T^2 f = T^2 f \circ \kappa_M$ for each $f \in C^k(M, N)$;
- (2) $T(\pi_M) \circ \kappa_M = \pi_{TM}$;
- (3) $\pi_{TM} \circ \kappa_M = T(\pi_M)$;
- (4) $\kappa_M^{-1} = \kappa_M$;
- (5) κ_M is a linear isomorphism from the bundle $(TTM, T(\pi_M), TM)$ to (TTM, π_{TM}, TM) , so it interchanges the two vector bundle structures on TTM ;
- (6) κ_M is the unique smooth map $TTM \rightarrow TTM$ which, for each $\gamma : \mathbb{R} \rightarrow M$, satisfies

$$\partial_t \partial_s \gamma(t, s) = \kappa_M \partial_t \partial_s \gamma(t, s).$$

In a similar way the second cotangent bundle of a manifold M can be defined. Even more, for every manifold there is a *geometrical isomorphism* between the bundles $TT^*M = T(T^*M)$ and $T^*TM = T^*(TM)$ [Modugno and Stefani (1978)].

4.3.2 The Natural Vector Bundle

In this section we mainly follow [Michor (2001); Kolar *et al.* (1993)].

A *vector bundle functor* or *natural vector bundle* is a functor \mathcal{F} which associates a vector bundle $(\mathcal{F}(M), \pi_M, M)$ to each n -manifold M and a vector bundle homomorphism

$$\begin{array}{ccc} \mathcal{F}(M) & \xrightarrow{\mathcal{F}(\varphi)} & \mathcal{F}(N) \\ \pi_M \downarrow & & \downarrow \pi_N \\ M & \xrightarrow{\varphi} & N \end{array}$$

to each $\varphi : M \rightarrow N$ in \mathcal{M} , which covers φ and is fiberwise a linear isomorphism. Two common examples of the vector bundle functor \mathcal{F} are tangent bundle functor T and cotangent bundle functor T^* (see section 3.5).

The space of all smooth sections of the vector bundle (E, π_M, M) is denoted by $\Gamma(E, \pi_M, M)$. Clearly, it is a vector space with fiberwise addition and scalar multiplication.

Let \mathcal{F} be a vector bundle functor on \mathcal{M} . Let M be a smooth manifold and let $X \in \mathcal{X}(M)$ be a vector-field on M . Then the flow F_t of X for fixed

t , is a diffeomorphism defined on an open subset of M . The map

$$\begin{array}{ccc} \mathcal{F}(M) & \xrightarrow{\mathcal{F}(F_t)} & \mathcal{F}(M) \\ \pi_M \downarrow & & \downarrow \pi_M \\ M & \xrightarrow{F_t} & M \end{array}$$

is then a vector bundle isomorphism, defined over an open subset of M .

We consider a tensor-field τ (3.6), which is a section $\tau \in \Gamma(\mathcal{F}(M))$ of the vector bundle $(\mathcal{F}(M), \pi_M, M)$ and we define for $t \in \mathbb{R}$

$$F_t^* \tau = \mathcal{F}(F_{-t}) \circ \tau \circ F_t,$$

a local section of the bundle $\mathcal{F}(M)$. For each point $m \in M$ the value $F_t^* \tau(x) \in \mathcal{F}(M)_m$ is defined, if t is small enough (depending on x). So, in the vector space $\mathcal{F}(M)_m$ the expression $\frac{d}{dt} \big|_{t=0} F_t^* \tau(x)$ makes sense and therefore the section

$$\mathcal{L}_X \tau = \frac{d}{dt} \big|_{t=0} F_t^* \tau$$

is globally defined and is an element of $\Gamma(\mathcal{F}(M))$. It is called the *Lie derivative of the tensor-field τ along a vector-field $X \in \mathcal{X}(M)$* (see section 3.7, for details on Lie derivative).

In this situation we have:

- (1) $F_t^* F_r^* \tau = F_{t+r}^* \tau$, whenever defined.
- (2) $\frac{d}{dt} F_t^* \tau = F_t^* \mathcal{L}_X \tau = \mathcal{L}_X (F_t^* \tau)$, so
 $[\mathcal{L}_X, F_t^*] = \mathcal{L}_X \circ F_t^* - F_t^* \circ \mathcal{L}_X = 0$, whenever defined.
- (3) $F_t^* \tau = \tau$ for all relevant t iff $\mathcal{L}_X \tau = 0$.

Let \mathcal{F}_1 and \mathcal{F}_2 be two vector bundle functors on \mathcal{M} . Then the (fiberwise) tensor product $(\mathcal{F}_1 \otimes \mathcal{F}_2)(M) = \mathcal{F}_1(M) \otimes \mathcal{F}_2(M)$ is again a vector bundle functor and for $\tau_i \in \Gamma(\mathcal{F}_i(M))$ with $i = 1, 2$, there is a section $\tau_1 \otimes \tau_2 \in \Gamma(\mathcal{F}_1 \otimes \mathcal{F}_2)(M)$, given by the pointwise tensor product.

Also in this situation, for $X \in \mathcal{X}(M)$ we have

$$\mathcal{L}_X (\tau_1 \otimes \tau_2) = \mathcal{L}_X \tau_1 \otimes \tau_2 + \tau_1 \otimes \mathcal{L}_X \tau_2.$$

In particular, for $f \in C^k(M, \mathbb{R})$ we have $\mathcal{L}_X (f \tau) = df(X) \tau + f \mathcal{L}_X \tau$.

For any vector bundle functor \mathcal{F} on \mathcal{M} and $X, Y \in \mathcal{X}(M)$ we have:

$$[\mathcal{L}_X, \mathcal{L}_Y] = \mathcal{L}_X \circ \mathcal{L}_Y - \mathcal{L}_Y \circ \mathcal{L}_X = \mathcal{L}_{[X, Y]} : \Gamma(\mathcal{F}(M)) \rightarrow \Gamma(\mathcal{F}(M)).$$

4.3.3 Vertical Tangent and Cotangent Bundles

4.3.3.1 Tangent and Cotangent Bundles Revisited

Recall (from section 3.5 above) that the most important vector bundles are familiar tangent and cotangent bundles. The fibres of the tangent bundle $\pi_M : TM \rightarrow M$ of a manifold M are tangent spaces to M . The peculiarity of the tangent bundle TM in comparison with other vector bundles over M lies in the fact that, given an atlas $\Psi_M = \{(U_\xi, \phi_\xi)\}$ of a manifold M , the tangent bundle of M admits the holonomic atlas $\Psi = \{(U_\xi, \psi_\xi = T\phi_\xi)\}$, where by $T\phi_\xi$ is denoted the tangent map to ϕ_ξ . Namely, given coordinates x^α on a manifold M , the associated bundle coordinates on TM are holonomic coordinates (\dot{x}^α) with respect to the holonomic frames $\{\partial_\alpha\}$ for tangent spaces $T_x M$, $x \in M$. Their transition functions read

$$\dot{x}'^\alpha = \frac{\partial x'^\alpha}{\partial x^\mu} \dot{x}^\mu.$$

Every manifold map $f : M \rightarrow M'$ induces the linear bundle map over f of the tangent bundles (4.2).

The cotangent bundle of a manifold M is the dual $\pi_{*M} : T^*M \rightarrow M$ of the tangent bundle $TM \rightarrow M$. It is equipped with the holonomic coordinates $(x^\alpha, \dot{x}_\alpha)$ with respect to the coframes $\{dx^\alpha\}$ for T^*M which are the duals of $\{\partial_\alpha\}$. Their transition functions read

$$\dot{x}'_\alpha = \frac{\partial x^\mu}{\partial x'^\alpha} \dot{x}_\mu.$$

Recall that a tensor product of tangent and cotangent bundles over M ,

$$T = (\otimes^m TM) \otimes (\otimes^k T^*M), \quad (m, k \in \mathbb{N}), \quad (4.11)$$

is called a *tensor bundle*. Given two vector bundles Y and Y' over the same base X , their tensor product $Y \otimes Y'$ is a vector bundle over X whose fibres are the tensor products of those of the vector bundles Y and Y' .

Tangent, cotangent and tensor bundles belong to the category BUN of natural fibre bundles which admit the canonical lift of any diffeomorphism f of a base to a bundle automorphism, called the natural automorphism [Kolar *et al.* (1993)]. For example, the natural automorphism of the tangent bundle TM over a diffeomorphism f of its base M is the tangent map Tf (4.2) over f . In view of the expression (4.2), natural automorphisms are also called holonomic transformations or general covariant transformations (in gravitation theory).

Let $TY \rightarrow Y$ be the tangent bundle of a bundle $Y \rightarrow X$. The following diagram commutes

$$\begin{array}{ccc} TY & \xrightarrow{T\pi} & TX \\ \pi_Y \downarrow & & \downarrow \pi_X \\ Y & \xrightarrow{\pi} & X \end{array}$$

where $T\pi : TY \rightarrow TX$ is a fibre bundle. Note that $T\pi$ is still the bundle map of the bundle $TY \rightarrow Y$ to TX over π and the fibred map of the bundle $TY \rightarrow X$ to TX over X . There is also the canonical surjection

$$\pi_T : TY \longrightarrow TX \longrightarrow Y, \quad \text{given by} \quad \pi_T = \pi_X \circ T\pi = \pi \circ \pi_Y.$$

Now, given the fibre coordinates (x^α, y^i) of a fibre bundle Y , the corresponding induced coordinates of TY are

$$(x^\alpha, y^i, \dot{x}^\alpha, \dot{y}^i), \quad \dot{y}^i = \frac{\partial y^i}{\partial y^j} \dot{y}^j.$$

This expression shows that the tangent bundle $TY \rightarrow Y$ of a fibre bundle Y has the vector subbundle

$$VY = \text{Ker } T\pi$$

where $T\pi$ is regarded as the fibred map of $TY \rightarrow X$ to TX over X . The subbundle VY consists of tangent vectors to fibres of Y . It is called the *vertical tangent bundle* of Y and provided with the induced coordinates $(x^\alpha, y^i, \dot{y}^i)$ with respect to the fibre bases $\{\partial_i\}$.

The *vertical cotangent bundle* $V^*Y \rightarrow Y$ of a fibre bundle $Y \rightarrow X$ is defined as the dual of the vertical tangent bundle $VY \rightarrow Y$. Note that it is not a subbundle of the cotangent bundle T^*Y , but there is the canonical surjection

$$\zeta : T^*Y \longrightarrow V^*Y, \quad \dot{x}_\alpha dx^\alpha + \dot{y}_i dy^i \mapsto \dot{y}_i \bar{d}y^i, \quad (4.12)$$

where $\{\bar{d}y^i\}$ are the bases for the fibres of V^*Y which are duals of the holonomic frames $\{\partial_i\}$ for the vertical tangent bundle VY .

With VY and V^*Y , we have the following short exact sequences of vector bundles over a fibre bundle $Y \rightarrow X$:

$$0 \rightarrow VY \hookrightarrow TY \rightarrow Y \times TX \rightarrow 0, \quad (4.13)$$

$$0 \rightarrow Y \times T^*X \hookrightarrow T^*Y \rightarrow V^*Y \rightarrow 0 \quad (4.14)$$

Every splitting

$$Y \times TX \hookrightarrow TY, \quad \partial_\alpha \mapsto \partial_\alpha + \Gamma_\alpha^i(y)\partial_i,$$

of the exact sequence (4.13) and

$$V^*Y \rightarrow T^*Y, \quad \bar{d}^i \mapsto dy^i - \Gamma_\alpha^i(y)dx^\alpha,$$

of the exact sequence (4.14), by definition, corresponds to a certain *connection* on the bundle $Y \rightarrow X$, and *vice versa*.

Let Φ be a fibred map of a bundle $Y \rightarrow X$ to a bundle $Y' \rightarrow X'$ over $f : X \rightarrow X'$. The tangent map $T\Phi : TY \rightarrow TY'$ to Φ reads

$$(\dot{x}'^\alpha, \dot{y}'^i) \circ T\Phi = (\partial_\mu f^\alpha \dot{x}^\mu, \partial_\mu \Phi^i \dot{x}^\mu + \partial_j \Phi^i \dot{y}^j). \quad (4.15)$$

It is both the linear bundle map over Φ , given by the commutativity diagram

$$\begin{array}{ccc} TY & \xrightarrow{T\Phi} & TY' \\ \pi_Y \downarrow & & \downarrow \pi_{Y'} \\ Y & \xrightarrow{\Phi} & Y' \end{array}$$

as well as the fibred map over the tangent map Tf to f , given by the commutativity diagram

$$\begin{array}{ccc} TY & \xrightarrow{T\Phi} & TY' \\ \downarrow & & \downarrow \\ TX & \xrightarrow{Tf} & TX' \end{array}$$

4.3.4 Affine Bundles

Given a vector bundle $\bar{Y} \rightarrow X$, an *affine bundle* modelled over \bar{Y} is a fibre bundle $Y \rightarrow X$ whose fibres Y_x , (for all $x \in X$), are affine spaces modelled over the corresponding fibres \bar{Y}_x of the vector bundle \bar{Y} , and Y admits a bundle atlas Ψ_Y (4.8) whose trivialization morphisms $\psi_\xi(x)$ and transition functions $\rho_{\xi\zeta}(x)$ are affine maps. The corresponding bundle coordinates (y^i) possess an affine coordinate transformation law

$$y'^i = \rho_j^i(x^\alpha) y^j + \rho^i(x^\alpha).$$

In other words, an affine bundle admits an atlas of affine bundle coordinates (x^α, y^i) such that

$$r : (x^\alpha, y^i) \times (x^\alpha, \bar{y}^i) \mapsto (x^\alpha, y^i + \bar{y}^i)$$

where (x^α, \bar{y}^i) are linear bundle coordinates of the vector bundle \bar{Y} . In particular, every vector bundle Y has the canonical structure of an affine bundle modelled on Y itself by the map

$$r : (y, y') \mapsto y + y'.$$

Every affine bundle has a global section.

One can define a direct sum $Y \oplus Y'$ of a vector bundle $\bar{Y}' \rightarrow X$ and an affine bundle $Y \rightarrow X$ modelled over a vector bundle $Y \rightarrow X$, as is an affine bundle modelled over the Whitney sum of vector bundles $Y' \oplus \bar{Y}$.

Affine bundles are subject to *affine bundle maps* which are affine fibre-wise maps. Any affine bundle map $\Phi : Y \rightarrow Y'$ from an affine bundle Y modelled over a vector bundle \bar{Y} to an affine bundle Y' modelled over a vector bundle \bar{Y}' , induces the linear bundle map of these vector bundles

$$\bar{\Phi} : \bar{Y} \rightarrow \bar{Y}', \quad \bar{y}'^i \circ \bar{\Phi} = \frac{\partial \Phi^i}{\partial y^j} \bar{y}^j. \quad (4.16)$$

4.4 Application: Semi-Riemannian Geometrical Mechanics

In this subsection we develop a *Finsler-like approach* to semi-Riemannian geometrical dynamics.

4.4.1 Vector-Fields and Connections

Let M be an n D smooth manifold. Recall that a smooth (C^∞) vector-field X on M defines the flow

$$\dot{x} = X(x). \quad (4.17)$$

By definition, a semi-Riemannian metric g on M is a smooth symmetric tensor-field of type $(0, 2)$ which assigns to each point $x \in M$ a nondegenerate inner product $g(x)$ on the tangent space $T_x M$ of signature (r, s) . The pair (M, g) is called a *semi-Riemannian manifold*.

The vector-field X and the semi-Riemannian metric g determine the *energy* $f : M \rightarrow \mathbb{R}$, given by $f = \frac{1}{2}g(X, X)$. The vector-field X (and its flow) on (M, g) is called [Udriste (2000)]:

- (1) *time-like*, if $f < 0$;
- (2) *nonspacelike or causal*, if $f \leq 0$;
- (3) *null or lightlike*, if $f = 0$;
- (4) *space-like*, if $f > 0$.

Let ∇ be the *Levi-Civita connection* of (M, g) . Using the semi-Riemannian version of the covariant derivative operator (3.157), we get the *prolongation*

$$\frac{\nabla}{dt} \dot{x} = \nabla_{\dot{x}} X \quad (4.18)$$

of the differential system (4.17) or of any perturbation of the system (4.17) get adding to the second member X a parallel vector-field Y with respect to the covariant derivative ∇ . The prolongation by derivation represents the general dynamics of the flow. The vector-field Y can be used to illustrate a progression from stable to unstable flows, or converse.

The vector-field X , the metric g , and the connection ∇ determine the external $(1, 1)$ -tensor-field

$$F = \nabla X - g^{-1} \otimes g(\nabla X), \quad F_j^i = \nabla_j X^i - g^{ih} g_{kj} \nabla_h X^k,$$

(with $i, j, h, k = 1, \dots, n$), which characterizes the *helicity* of vector-field X and its flow.

First we write the differential system (4.18) in the equivalent form

$$\frac{\nabla}{dt} \dot{x} = g^{-1} \otimes g(\nabla X)(\dot{x}) + F(\dot{x}). \quad (4.19)$$

Successively we modify the differential system (4.19) as follows [Udriste (2000)]:

$$\frac{\nabla}{dt} \dot{x} = g^{-1} \otimes g(\nabla X)(X) + F(\dot{x}), \quad (4.20)$$

$$\frac{\nabla}{dt} \dot{x} = g^{-1} \otimes g(\nabla X)(\dot{x}) + F(X), \quad (4.21)$$

$$\frac{\nabla}{dt} \dot{x} = g^{-1} \otimes g(\nabla X)(X) + F(X). \quad (4.22)$$

Obviously, the second-order systems (4.20), (4.21), (4.22) are prolongations of the first-order system (4.17). Each of them is connected either to the dynamics of the field X or to the dynamics of a particle which is sensitive to the vector-field X . Since

$$g^{-1} \otimes g(\nabla X)(X) = \text{grad } f,$$

we shall show that the prolongation (4.20) describes a conservative dynamics of the vector-field X or of a particle which is sensitive to the vector-field X . The physical phenomenon produced by (4.21) or (4.22) was not yet studied [Udriste (2000)].

In the case $F = 0$, the kinematic system (4.17) prolongs to a potential dynamical system with n degrees of freedom, namely

$$\frac{\nabla}{dt}\dot{x} = \text{grad } f. \quad (4.23)$$

In the case $F \neq 0$, the kinematic system (4.17) prolongs to a non-potential dynamical system with n degrees of freedom, namely

$$\frac{\nabla}{dt}\dot{x} = \text{grad } f + F(\dot{x}). \quad (4.24)$$

Let us show that the dynamical systems (4.23) and (4.24) are conservative. To simplify the exposition we identify the tangent bundle TM with the cotangent bundle T^*M using the semi-Riemann metric g [Udriste (2000)]. The trajectories of the dynamical system (4.23) are the extremals of the Lagrangian

$$L = \frac{1}{2}g(\dot{x}, \dot{x}) + f(x).$$

The trajectories of the dynamical system (4.24) are the extremals of the Lagrangian

$$L = \frac{1}{2}g(\dot{x} - X, \dot{x} - X) = \frac{1}{2}g(\dot{x}, \dot{x}) - g(X, \dot{x}) + f(x).$$

The dynamical systems (4.23) and (4.24) are conservative, the Hamiltonian being the same for both cases, namely

$$H = \frac{1}{2}g(\dot{x}, \dot{x}) - f(x).$$

The restriction of the Hamiltonian H to the flow of the vector-field X is zero.

4.4.2 Hamiltonian Structures on the Tangent Bundle

Let (N, ω) be a $2n$ D symplectic phase-space manifold, and $H : N \rightarrow \mathbb{R}$ be a C^∞ real function. We define the Hamiltonian gradient X_H as being the vector-field which satisfies

$$\omega_p(X_H(x), v) = dH(x)(v), \quad (\text{for all } v \in T_x N),$$

and the Hamiltonian equations as

$$\dot{x} = X_H(x).$$

Let (M, g) be a semi-Riemann n D manifold. Let X be a C^∞ vector-field on M , and $\omega = g \circ F$ the two-form associated to the tensor-field $F = \nabla X - g^{-1} \otimes g(\nabla X)$ via the metric g .

The tangent bundle is usually equipped with the *Sasakian metric* G , induced by g ,

$$G = g_{ij}dx^i \otimes dx^j + g_{ij}\delta y^i \otimes \delta y^j.$$

If (x^i, y^i) are the coordinates of the point $(x, y) \in TM$ and Γ_{jk}^i are the components of the connection induced by g_{ij} , then we have the following *dual frames* [Udriste (2000)]

$$\left(\frac{\delta}{\delta x^i} = \frac{\partial}{\partial x^i} - \Gamma_{ij}^h y^j \frac{\partial}{\partial y^h}, \quad \frac{\partial}{\partial y^i} \right) \subset \mathcal{X}(TM), \quad \text{and}$$

$$(dx^j, \quad \delta y^j = dy^j + \Gamma_{hk}^j y^k dx^h) \subset \mathcal{X}^*(TM).$$

The dynamical system (4.23) lifts to TM as a Hamiltonian dynamical system with respect to the

$$\text{Hamiltonian} \quad H = \frac{1}{2}g(\dot{x}, \dot{x}) - f(x), \quad \text{and}$$

$$\text{symplectic two-form} \quad \Omega_1 = g_{ij}dx^i \omega \delta y^j.$$

This can be verified by putting $\eta_1 = g_{ij}y^i dx^j$, and $d\eta_1 = -\Omega_1$.

The dynamical system (4.24) lifts to TM as a Hamiltonian dynamical system with respect to the above Hamiltonian function and the symplectic two-form

$$\Omega_2 = \frac{1}{2}\omega_{ij}dx^i \omega dx^j + g_{ij}dx^i \omega \delta y^j.$$

This can be verified by putting $\eta_2 = -g_{ij}X^i dx^j + g_{ij}y^i dx^j$, and $d\eta_2 = -\Omega_2$.

In the remainder of this subsection, we give three examples in Euclidean spaces, so we can put all indices down (still summing over repeated indices).

Pendulum Geometry

We use the Riemannian manifold $(\mathbb{R}^2, \delta_{ij})$. The small oscillations of a plane pendulum are described as solutions of the following differential system giving the plane pendulum flow,

$$\dot{x}_1 = -x_2, \quad \dot{x}_2 = x_1. \quad (4.25)$$

In this case, the set $\{x_1(t) = 0, x_2(t) = 0, (t \in \mathbb{R})\}$ is the equilibrium point and

$$x_1(t) = c_1 \cos t + c_2 \sin t, \quad x_2(t) = c_1 \sin t - c_2 \cos t$$

is the general solution, which is a family of circles with a common center.

$$\begin{aligned} \text{Let } X &= (X_1, X_2), & X_1(x_1, x_2) &= -x_2, & X_2(x_1, x_2) &= x_1, \\ f(x_1, x_2) &= \frac{1}{2}(x_1^2 + x_2^2), & \text{curl } X &= (0, 0, 2), & \text{div } X &= 0. \end{aligned}$$

The pendulum flow conserves the areas. The prolongation by derivation of the kinematic system (4.25) is [Udriste (2000)]

$$\begin{aligned} \ddot{x}_i &= \frac{\partial X_i}{\partial x_j} \dot{x}_j, & (i, j = 1, 2) \\ \text{or } \ddot{x}_1 &= -\dot{x}_2, & \ddot{x}_2 = \dot{x}_1. \end{aligned}$$

This prolongation admits a family of circles as the general solution

$$x_1(t) = a_1 \cos t + a_2 \sin t + h, \quad x_2(t) = a_1 \sin t - a_2 \cos t + k, \quad (t \in \mathbb{R}).$$

The pendulum geometrodynamics is described by

$$\begin{aligned} \ddot{x}_i &= \frac{\partial f}{\partial x_i} + \left(\frac{\partial X_i}{\partial x_j} - \frac{\partial X_j}{\partial x_i} \right) \dot{x}_j, & (i, j = 1, 2), \\ \text{or } \ddot{x}_1 &= x_1 - 2\dot{x}_2, & \ddot{x}_2 = x_2 + 2\dot{x}_1, \end{aligned} \quad (4.26)$$

with a family of spirals as the general solution

$$\begin{aligned} x_1(t) &= b_1 \cos t + b_2 \sin t + b_3 t \cos t + b_4 t \sin t, \\ x_2(t) &= b_1 \sin t - b_2 \cos t + b_3 t \sin t - b_4 t \cos t, & (t \in \mathbb{R}). \end{aligned}$$

Using

$$\begin{aligned} L &= \frac{1}{2} \left[(\dot{x}_1)^2 + (\dot{x}_2)^2 \right] + x_2 \dot{x}_1 - x_1 \dot{x}_2 + f, \\ H &= \frac{1}{2} \left[(\dot{x}_1)^2 + (\dot{x}_2)^2 \right] - f, \quad g_{ij} = (H + f) \delta_{ij}, \\ N_j^i &= -F_j^i = -\delta^{ih} F_{jh}, \quad F_{ij} = \frac{\partial X_j}{\partial x_i} - \frac{\partial X_i}{\partial x_j}, \quad (i, j, h = 1, 2), \end{aligned}$$

the solutions of the differential system (4.26) are horizontal pregeodesics of the Riemann–Jacobi–Lagrangian manifold $(\mathbb{R}^2 \setminus \{0\}, g_{ij}, N_j^i)$.

Geometry of the Lorenz Flow

We use the Riemannian manifold $(\mathbb{R}^3, \delta_{ij})$. The *Lorenz flow* is a first dissipative model with *chaotic behavior* discovered in numerical experiment. Its state equations are (see [Lorenz (1963); Sparrow (1982)])

$$\dot{x}_1 = -\sigma x_1 + \sigma x_2, \quad \dot{x}_2 = -x_1 x_3 + r x_1 - x_2, \quad \dot{x}_3 = x_1 x_2 - b x_3,$$

where σ, r, b are real parameters. Usually σ, b are kept fixed whereas r is varied. At

$$r > r_0 = \frac{\sigma(\sigma + b + 3)}{\sigma - b - 1}$$

chaotic behavior is observed [Udriste (2000)].

$$\begin{aligned} \text{Let } X &= (X_1, X_2, X_3), \quad X_1(x_1, x_2, x_3) = -\sigma x_1 + \sigma x_2, \\ X_2(x_1, x_2, x_3) &= -x_1 x_3 + r x_1 - x_2, \quad X_3(x_1, x_2, x_3) = x_1 x_2 - b x_3, \\ f &= \frac{1}{2} [(-\sigma x_1 + \sigma x_2)^2 + (-x_1 x_3 + r x_1 - x_2)^2 + (x_1 x_2 - b x_3)^2], \\ \text{curl } X &= (2x_1, -x_2, r - x_3 - \sigma). \end{aligned}$$

The *Lorenz dynamics* is described by

$$\begin{aligned} \ddot{x}_i &= \frac{\partial f}{\partial x_i} + \left(\frac{\partial X_i}{\partial x_j} - \frac{\partial X_j}{\partial x_i} \right) \dot{x}_j, \quad (i, j = 1, 2, 3), \quad \text{or} \\ \ddot{x}_1 &= \frac{\partial f}{\partial x_1} + (\sigma + x_3 - r) \dot{x}_2 - x_2 \dot{x}_3, \quad \ddot{x}_2 = \frac{\partial f}{\partial x_2} + (r - x_3 - \sigma) \dot{x}_1 - 2x_1 \dot{x}_3, \\ \ddot{x}_3 &= \frac{\partial f}{\partial x_3} + x_2 \dot{x}_1 + 2x_1 \dot{x}_2. \end{aligned} \tag{4.27}$$

$$\text{Using} \quad L = \frac{1}{2} \sum_{i=1}^3 (\dot{x}_i)^2 - \sum_{i=1}^3 X_i \dot{x}_i + f, \quad H = \frac{1}{2} \sum_{i=1}^3 (\dot{x}_i)^2 - f,$$

$$g_{ij} = (H + f)\delta_{ij}, \quad N_j^i = -F_j^i = -\delta^{ih} F_{jh}, \quad F_{ij} = \frac{\partial X_j}{\partial x_i} - \frac{\partial X_i}{\partial x_j},$$

$(i, j, h = 1, 2, 3)$, the solutions of the differential system (4.27) are horizontal pregeodesics of the Riemann–Jacobi–Lagrangian manifold $(\mathbb{R}^3 \setminus E, g_{ij}, N_j^i)$, where E is the set of equilibrium points.

Geometry of the ABC Flow

We use the Riemannian manifold $(\mathbb{R}^3, \delta_{ij})$. One example of a fluid velocity that contains exponential stretching and hence instability is the ABC flow, named after Arnold, Beltrami and Childress,

$$\dot{x}_1 = A \sin x_3 + C \cos x_2, \quad \dot{x}_2 = B \sin x_1 + A \cos x_3, \quad \dot{x}_3 = C \sin x_2 + B \cos x_1.$$

For nonzero values of the constants A, B, C the preceding system is not globally integrable. The topology of the flow lines is very complicated and can only be investigated numerically to reveal regions of chaotic behavior. The ABC flow conserves the volumes since the ABC field is solenoidal.

The ABC geometrodynamics is described by [Udriste (2000)]

$$\ddot{x}_i = \frac{\partial f}{\partial x_i} + \left(\frac{\partial X_i}{\partial x_j} - \frac{\partial X_j}{\partial x_i} \right) \dot{x}_j, \quad (i, j = 1, 2, 3).$$

Since $f = \frac{1}{2}(A + B + C + 2AC \sin x_3 \cos x_2 + 2BA \sin x_1 \cos x_3 + 2CB \sin x_2 \cos x_1)$, and $\text{curl } X = X$, the ABC geometrodynamics is given by the system,

$$\begin{aligned} \ddot{x}_1 &= AB \cos x_1 \cos x_3 - BC \sin x_1 \sin x_2 \\ &\quad - (B \cos x_1 + C \sin x_2) \dot{x}_2 + (B \sin x_1 + A \cos x_3) \dot{x}_3, \\ \ddot{x}_2 &= -AC \sin x_2 \sin x_3 + BC \cos x_1 \cos x_2 \\ &\quad + (B \cos x_1 + C \sin x_2) \dot{x}_1 - (A \sin x_3 + C \cos x_2) \dot{x}_2, \\ \ddot{x}_3 &= AC \cos x_3 \cos x_2 - BA \sin x_1 \sin x_3 \\ &\quad - (B \sin x_1 + A \cos x_3) \dot{x}_1 + (C \cos x_2 + A \sin x_3) \dot{x}_2. \end{aligned}$$

Using

$$L = \frac{1}{2} \dot{x}_i \dot{x}_i - X_i \dot{x}_i + f, \quad H = \frac{1}{2} \dot{x}_i \dot{x}_i - f, \quad (i, j, h = 1, 2, 3),$$

$$g_{ij} = (H + f) \delta_{ij}, \quad N_j^i = -F_j^i = -\delta^{ih} F_{jh}, \quad F_{ij} = \frac{\partial X_j}{\partial x_i} - \frac{\partial X_i}{\partial x_j},$$

the solutions of the above differential system are horizontal pregeodesics of the Riemann–Jacobi–Lagrangian manifold $(\mathbb{R}^3 \setminus E, g_{ij}, N_j^i)$, where E is the set of equilibrium points, which is included in the surface of equation

$$\sin x_1 \sin x_2 \sin x_3 + \cos x_1 \cos x_2 \cos x_3 = 0.$$

4.5 K –Theory and Its Applications

Recall from [Dieudonne (1988)] that the 1930s were the decade of the development of the *cohomology theory*, as several research directions grew together and the *de Rham cohomology*, that was implicit in Poincaré’s work, became the subject of definite theorems. The development of algebraic topology from 1940 to 1960 was very rapid, and the role of homology theory was often as ‘baseline’ theory, easy to compute and in terms of which topologists sought to calculate with other functors. The axiomatization of homology theory by Eilenberg and Steenrod (celebrated *Eilenberg–Steenrod Axioms*) revealed that what various candidate homology theories had in common was, roughly speaking, some *exact sequences* (in particular, the *Mayer–Vietoris Theorem* and the *Dimension Axiom* that calculated the homology of the point).

4.5.1 Topological K –Theory

Now, K –theory is an *extraordinary cohomology theory*, which consists of topological K –theory and algebraic K –theory. The *topological K –theory* was founded to study vector bundles on general topological spaces, by means of ideas now recognisee as (general) K –theory that were introduced by Alexander Grothendieck. The early work on topological K –theory was due to Michael Atiyah and Friedrich Hirzebruch.

Let X be a compact Hausdorff space and $k = \mathbb{R}$ or $k = \mathbb{C}$. Then $K_k(X)$ is the *Grothendieck group of the commutative monoid*² which elements are

²Recall that a monoid is an algebraic structure with a single, associative binary operation and an identity element; a monoid whose operation is commutative is called a commutative monoid (or, an Abelian monoid); e.g., every group is a monoid and every

the isomorphism classes of finite dimensional k -vector bundles on X with the operation

$$[E \oplus F] := [E] \oplus [F]$$

for vector bundles E, F .³ Usually, $K_k(X)$ is denoted $KO(X)$ in real case and $KU(X)$ in the complex case.

More precisely, the *stable equivalence*, i.e., the equivalence relation on bundles E and F on X of defining the same element in $K(X)$, occurs when there is a trivial bundle G , so that $E \oplus G \cong F \oplus G$. Under the *tensor product of vector bundles*, $K(X)$ then becomes a *commutative ring*. The rank of a vector bundle carries over to the K -group define the homomorphism: $K(X) \rightarrow \check{H}^0(X, \mathbb{Z})$, where $\check{H}^0(X, \mathbb{Z})$ is the 0-group of the *Cech cohomology* which is equal to group of locally constant functions with values in \mathbb{Z} .

The constant map $X \rightarrow \{x_0\}$, $x_0 \in X$ defines the reduced K -group (of reduced homology)

$$\tilde{K}(X) = \text{Coker}(K(X) \rightarrow \{x_0\}).$$

In particular, when X is a connected space, then

$$\tilde{K}(X) \cong \text{Ker}(K(X) \rightarrow \check{H}^0(X, \mathbb{Z}) = \mathbb{Z}).$$

4.5.1.1 Bott Periodicity Theorem

An important property in the topological K -theory is the *Bott Periodicity Theorem* [Bott (1959)]⁴, which can be formulated this way:

Abelian group a commutative monoid.

³The Grothendieck group construction in abstract algebra constructs an Abelian group from a commutative monoid ‘in the best possible way’.

⁴The Bott Periodicity Theorem is a result from homotopy theory discovered by Raoul Bott during the latter part of the 1950s, which proved to be of foundational significance for much further research, in particular in K -theory of stable complex vector bundles, as well as the stable homotopy groups of spheres. Bott periodicity can be formulated in numerous ways, with the periodicity in question always appearing as a period 2 phenomenon, with respect to dimension, for the theory associated to the unitary group. The context of Bott periodicity is that the homotopy groups of spheres, which would be expected to play the basic part in algebraic topology by analogy with homology theory, have proved elusive (and the theory is complicated). The subject of stable homotopy theory was conceived as a simplification, by introducing the suspension (smash product with a circle) operation, and seeing what (roughly speaking) remained of homotopy theory once one was allowed to suspend both sides of an equation, as many times as one wished. The stable theory was still hard to compute with, in practice. What Bott periodicity offered was an insight into some highly non-trivial spaces, with central status

- (1) $K(X \times S^2) = K(X) \otimes K(S^2)$, and $K(S^2) = [H]/(H-1)^2$, where H is the class of the *tautological bundle* on the $S^2 = P^1$, i.e., the *Riemann sphere* as complex projective line;
- (2) $\tilde{K}^{n+2}(X) = \tilde{K}^n(X)$;
- (3) $\Omega^2 \text{BU} \simeq \text{BU} \times \mathbb{Z}$.

In real K -theory there is a similar periodicity, but modulo 8.

4.5.2 Algebraic K -Theory

On the other hand, the so-called *algebraic K -theory* is an advanced part of *homological algebra* concerned with defining and applying a sequence $K_n(R)$ of functors from rings to Abelian groups, for $n = 0, 1, 2, \dots$. Here, for traditional reasons, the cases of K_0 and K_1 are thought of in somewhat different terms from the higher algebraic K -groups K_n for $n \geq 2$. In fact K_0 generalizes the construction of the ideal class group, using projective modules; and K_1 as applied to a commutative ring is the unit group construction, which was generalized to all rings for the needs of topology (simple homotopy theory) by means of elementary matrix theory. Therefore the first two cases counted as relatively accessible; while after that the theory becomes quite noticeably deeper, and certainly quite hard to compute (even when R is the ring of integers).

Historically, the roots of the theory were in *topological K -theory* (based on vector bundle theory); and its motivation the conjecture of Serre⁵ that now is the *Quillen–Suslin Theorem*.⁶

Applications of K -groups were found from 1960 onwards in *surgery*

in topology because of the connection of their cohomology with characteristic classes, for which all the (unstable) homotopy groups could be calculated. These spaces are the (infinite, or stable) unitary, orthogonal and symplectic groups U, O and Sp .

⁵Jean–Pierre Serre used the analogy of vector bundles with projective modules to found in 1959 what became algebraic K -theory. He formulated the *Serre’s Conjecture*, that projective modules over the ring of polynomials over a field are free modules; this resisted proof for 20 years.

⁶The Quillen–Suslin Theorem is a Theorem in abstract algebra about the relationship between free modules and projective modules. Projective modules are modules that are locally free. Not all projective modules are free, but in the mid-1950s, Jean–Pierre Serre found evidence that a limited converse might hold. He asked the question: Is every projective module over a polynomial ring over a field a free module? A more geometric variant of this question is whether every algebraic vector bundle on affine space is trivial. This was open until 1976, when Daniel Quillen and Andrei Suslin independently proved that the answer is yes. Quillen was awarded the Fields Medal in 1978 in part for his proof.

theory for manifolds, in particular; and numerous other connections with classical algebraic problems were found. A little later a branch of the theory for *operator algebras* was fruitfully developed. It also became clear that K -theory could play a role in *algebraic cycle theory* in algebraic geometry: here the higher K -groups become connected with the higher codimension phenomena, which are exactly those that are harder to access. The problem was that the definitions were lacking (or, too many and not obviously consistent). A definition of K_2 for fields by John Milnor, for example, gave an attractive theory that was too limited in scope, constructed as a quotient of the multiplicative group of the field tensored with itself, with some explicit relations imposed; and closely connected with central extensions [Milnor and Stasheff (1974)].

Eventually the foundational difficulties were resolved (leaving a deep and difficult theory), by a definition of D. Quillen:

$$K_n(R) = \pi_n(BGL(R)^+).$$

This is a very compressed piece of abstract mathematics. Here π_n is an n th homotopy group, $GL(R)$ is the direct limit of the general linear groups over R for the size of the matrix tending to infinity, B is the classifying space construction of homotopy theory, and the $+$ is *Quillen's plus construction*.

4.5.3 Chern Classes and Chern Character

An important properties in K -theory are the *Chern classes* and *Chern character* [Chern (1946)]. The Chern classes are a particular type of *characteristic classes* (topological invariants, see [Milnor and Stasheff (1974)]). associated to complex vector bundles of a smooth manifold. Recall that a characteristic class is a way of associating to each principal bundle on a topological space X a cohomology class of X . The cohomology class measures the extent to which the bundle is 'twisted' – particularly, whether it possesses sections or not. In other words, characteristic classes are global invariants which measure the deviation of a local product structure from a global product structure. They are one of the unifying geometric concepts in algebraic topology, differential geometry and algebraic geometry.⁷

⁷Recall that characteristic classes are in an essential way phenomena of cohomology theory – they are contravariant functors, in the way that a section is a kind of function on a space, and to lead to a contradiction from the existence of a section we do need that variance. In fact cohomology theory grew up after homology and homotopy theory, which are both covariant theories based on mapping into a space; and characteristic class theory in its infancy in the 1930s (as part of obstruction theory) was one major

If we describe the same vector bundle on a manifold in two different ways, the Chern classes will be the same, i.e., if the Chern classes of a pair of vector bundles do not agree, then the vector bundles are different (the converse is not true, though). In topology, differential geometry, and algebraic geometry, it is often important to count how many linearly independent sections a vector bundle has. The Chern classes offer some information about this through, for instance, the Riemann–Roch Theorem and the Atiyah–Singer Index Theorem. Chern classes are also feasible to calculate in practice. In differential geometry (and some types of algebraic geometry), the Chern classes can be expressed as polynomials in the coefficients of the *curvature form*.

In particular, given a complex hermitian vector bundle V of complex rank n over a smooth manifold M , a representative of each Chern class (also called a *Chern form*) $c_k(V)$ of V are given as the coefficients of the characteristic polynomial

$$\det \left(\frac{i t \Omega}{2\pi} + I \right) = c_k(V) t^k,$$

reason why a ‘dual’ theory to homology was sought. The characteristic class approach to curvature invariants was a particular reason to make a theory, to prove a general *Gauss–Bonnet Theorem*.

When the theory was put on an organized basis around 1950 (with the definitions reduced to homotopy theory) it became clear that the most fundamental characteristic classes known at that time (the *Stiefel–Whitney class*, the *Chern class*, and the *Pontryagin class*) were reflections of the classical linear groups and their maximal torus structure. What is more, the Chern class itself was not so new, having been reflected in the Schubert calculus on Grassmannians, and the work of the Italian school of algebraic geometry. On the other hand there was now a framework which produced families of classes, whenever there was a vector bundle involved.

The prime mechanism then appeared to be this: Given a space X carrying a vector bundle, implied in the homotopy category a mapping from X to a classifying space BG , for the relevant linear group G . For the homotopy theory, the relevant information is carried by compact subgroups such as the orthogonal groups and unitary groups of G . Once the cohomology $H^*(BG)$ was calculated, once and for all, the contravariance property of cohomology meant that characteristic classes for the bundle would be defined in $H^*(X)$ in the same dimensions. For example, the Chern class is really one class with graded components in each even dimension.

This is still the classic explanation, though in a given geometric theory it is profitable to take extra structure into account. When cohomology became ‘extra-ordinary’ with the arrival of K -theory and Thom’s cobordism theory from 1955 onwards, it was really only necessary to change the letter H everywhere to say what the characteristic classes were.

Characteristic classes were later found for foliations of manifolds; they have (in a modified sense, for foliations with some allowed singularities) a classifying space theory in homotopy theory.

of the curvature form Ω of V , which is defined as

$$\Omega = d\omega + \frac{1}{2}[\omega, \omega],$$

with ω the connection form and d the exterior derivative, or via the same expression in which ω is a *gauge form* for the *gauge group* of V . The scalar t is used here only as an indeterminate to generate the sum from the determinant, and I denotes the $n \times n$ identity matrix. To say that the expression given is a *representative* of the Chern class indicates that ‘class’ here means up to addition of an exact differential form. That is, Chern classes are cohomology classes in the sense of *de Rham cohomology*. It can be shown that the cohomology class of the Chern forms do not depend on the choice of connection in V .

For example, let \mathbb{CP}^1 be the *Riemann sphere*: a 1D complex projective space. Suppose that z is a holomorphic local coordinate for the Riemann sphere. Let $V = T\mathbb{CP}^1$ be the bundle of complex tangent vectors having the form $a\partial/\partial z$ at each point, where a is a complex number. In the following we prove the complex version of the *Hairy Ball Theorem*: *V has no section which is everywhere nonzero.*

For this, we need the following fact: the first Chern class of a trivial bundle is zero, i.e., $c_1(\mathbb{CP}^1 \times \mathbb{C}) = 0$. This is evinced by the fact that a trivial bundle always admits a flat metric. So, we will show that $c_1(V) \neq 0$.

Consider the Kähler metric

$$h = \frac{dzd\bar{z}}{(1 + |z|^2)}.$$

One can show that the curvature 2-form is given by

$$\Omega = \frac{2dz \wedge d\bar{z}}{(1 + |z|^2)^2}.$$

Furthermore, by the definition of the first Chern class

$$c_1 = \frac{i}{2\pi}\Omega.$$

We need to show that the cohomology class of this is non-zero. It suffices to compute its integral over the Riemann sphere:

$$\int c_1 = \frac{i}{\pi} \int \frac{dz \wedge d\bar{z}}{(1 + |z|^2)^2} = 2,$$

after switching to polar coordinates. By *Stokes Theorem*, an exact form would integrate to 0, so the cohomology class is nonzero. This proves that $T\mathbb{C}P^1$ is not a trivial vector bundle.

An important special case occurs when V is a *line bundle*. Then the only nontrivial Chern class is the *first Chern class*, which is an element of $H^2(X; \mathbb{Z})$ —the second cohomology group of X . As it is the top Chern class, it equals the *Euler class* of the bundle. The first Chern class turns out to be a complete topological invariant with which to classify complex line bundles. That is, there is a bijection between the isomorphism classes of line bundles over X and the elements of $H^2(X; \mathbb{Z})$, which associates to a line bundle its first Chern class. Addition in the second cohomology group coincides with tensor product of complex line bundles.

In algebraic geometry, this classification of (isomorphism classes of) complex line bundles by the first Chern class is a crude approximation to the classification of (isomorphism classes of) holomorphic line bundles by linear equivalence classes of divisors. For complex vector bundles of dimension greater than one, the Chern classes are not a complete invariant.

The Chern classes can be used to construct a homomorphism of rings from the topological K –theory of a space to (the completion of) its rational cohomology. For line bundles V , the Chern character ch is defined by

$$\text{ch}(V) = \exp[c_1(V)].$$

For sums of line bundles, the Chern character is defined by additivity. For arbitrary vector bundles, it is defined by pretending that the bundle is a sum of line bundles; more precisely, for sums of line bundles the Chern character can be expressed in terms of Chern classes, and we use the same formulas to define it on all vector bundles. For example, the first few terms are

$$\text{ch}(V) = \dim(V) + c_1(V) + c_1(V)^2/2 - c_2(V) + \dots$$

If V is filtered by line bundles L_1, \dots, L_k having first Chern classes x_1, \dots, x_k , respectively, then

$$\text{ch}(V) = e^{x_1} + \dots + e^{x_k} \dots$$

If a connection is used to define the Chern classes, then the explicit form of the Chern character is

$$\text{ch}(V) = \text{Tr} \left[\exp \left(\frac{i\Omega}{2\pi} \right) \right],$$

where Ω is the curvature of the connection.

The Chern character is useful in part because it facilitates the computation of the Chern class of a tensor product. Specifically, it obeys the following identities:

$$\mathrm{ch}(V \oplus W) = \mathrm{ch}(V) + \mathrm{ch}(W), \quad \mathrm{ch}(V \otimes W) = \mathrm{ch}(V)\mathrm{ch}(W).$$

Using the *Grothendieck Additivity Axiom* for Chern classes, the first of these identities can be generalized to state that ch is a homomorphism of Abelian groups from the K -theory $K(X)$ into the rational cohomology of X . The second identity establishes the fact that this homomorphism also respects products in $K(X)$, and so ch is a homomorphism of rings. The Chern character is used in the Hirzebruch–Riemann–Roch Theorem.

The so-called *twisted K -theory* a particular variant of K -theory, in which the twist is given by an integral 3D cohomology class. In physics, it has been conjectured to classify D -branes, Ramond–Ramond field strengths and in some cases even spinors in type II string theory.

4.5.4 Atiyah's View on K -Theory

According to Michael Atiyah [Atiyah and Anderson (1967); Atiyah (2000)], K -theory may roughly be described as the study of additive (or, Abelian) invariants of large matrices. The key point is that, although matrix multiplication is not commutative, matrices which act in orthogonal subspaces do commute. Given ‘enough room’ we can put matrices A and B into the block form

$$\begin{pmatrix} A & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & B \end{pmatrix},$$

which obviously commute. Examples of Abelian invariants are traces and determinants.

The prime motivation for the birth of K -theory came from Hirzebruch's generalization of the classical *Riemann–Roch Theorem* (see [Hirzebruch (1966)]). This concerns a complex projective algebraic manifold X and a holomorphic (or algebraic) vector bundle E over X . Then one has the sheaf cohomology groups $H^q(X, E)$, which are finite-dimensional vector spaces, and the corresponding *Euler characteristics*

$$\chi(X, E) = \sum_{q=0}^n (-1)^q \dim H^q(X, E),$$

where n is the complex dimension of X . One also has topological invariants of E and of the tangent bundle of X , namely their *Chern classes*. From these one defines a certain explicit polynomial $T(X, E)$ which by evaluation on X becomes a rational number. *Hirzebruch's Riemann–Roch Theorem* asserts the equality: $\chi(X, E) = T(X, E)$.

It is an important fact, easily proved, that both χ and T are additive for exact sequences of vector bundles:

$$0 \rightarrow E' \rightarrow E \rightarrow E'' \rightarrow 0,$$

$$\chi(X, E) = \chi(X, E') + \chi(X, E''), \quad T(X, E) = T(X, E') + T(X, E'').$$

This was the starting point of the Grothendieck's generalization. Grothendieck defined an Abelian group $K(X)$ as the universal additive invariant of exact sequences of algebraic vector bundles over X , so that χ and T both gave homomorphisms of $K(X)$ into the integers (or rationals).

More precisely, Grothendieck defined two different K -groups, one arising from vector bundles (denoted by K^0) and the other using coherent sheaves (denoted by K_0). These are formally analogous to cohomology and homology respectively. Thus $K^0(X)$ is a ring (under tensor product) while $K_0(X)$ is a $K^0(X)$ -module. Moreover, K^0 is contravariant while K_0 is covariant (using a generalization of χ). Finally, Grothendieck established the analogue of *Poincaré duality*. While $K^0(X)$ and $K_0(X)$ can be defined for an arbitrary projective variety X , singular or not, the natural map $K^0(X) \rightarrow K_0(X)$ is an isomorphism if X is non-singular.

The *Grothendieck's Riemann–Roch Theorem* concerns a morphism $f : X \rightarrow Y$ and compares the direct image of f in K -theory and cohomology. It reduces to the Hirzebruch's version when Y is a point.

Topological K-theory started with the famous *Bott Periodicity Theorem* [Bott (1959)], concerning the homotopy of the large unitary groups $U(N)$ (for $N \rightarrow \infty$). Combining Bott's Theorem with the formalism of Grothendieck, Atiyah and Hirzebruch, in the late 1950's, developed a K -theory based on topological vector bundles over a compact space [Atiyah and Hirzebruch (1961)]. Here, in addition to a group $K^0(X)$, they also introduced an odd counterpart $K^1(X)$, defined as the group of homotopy classes of X into $U(N)$, for N large. Putting these together,

$$K^*(X) = K^0(X) \oplus K^1(X),$$

they obtained a periodic 'generalized cohomology theory'. Over the ratio-

nals, the *Chern character* gave an isomorphism:

$$\text{ch} : K^*(X) \otimes Q \cong H^*(X, Q).$$

But, over the integers, K -theory is much more subtle and it has had many interesting topological applications. Most notable was the solution of the vector fields on spheres problem by Frank Adams, using real K -theory (based on the orthogonal groups) [Adams (1962)].

An old generalisation of K -theory is related to *projective bundles* [Atiyah and Anderson (1967); Atiyah (2000)]. Given a vector bundle V over a space X , we can form the bundle $P(V)$ whose fibre at $x \in X$ is the projective space $P(V_x)$. In terms of groups and principal bundles, this is the passage from $GL(n, C)$ to $PGL(n, C)$, or from $U(n)$ to $PU(n)$. We have two exact sequences of groups:

$$\begin{array}{ccccccc} 1 & \rightarrow & U(1) & \rightarrow & U(n) & \rightarrow & PU(n) \rightarrow 1 \\ 1 & \rightarrow & Z_n & \rightarrow & SU(n) & \rightarrow & PU(n) \rightarrow 1 \end{array}$$

The first gives rise to an obstruction $\alpha \in H^3(X, Z)$ to lifting a projective bundle to a vector bundle, while the second gives an obstruction $\beta \in H^2(X, Z_n)$ to lifting a projective bundle to a special unitary bundle. They are related by

$$\alpha = \delta(\beta), \quad \text{where} \quad \delta : H^2(X, Z_n) \rightarrow H^3(X, Z)$$

is the *coboundary operator*. This shows that $n\alpha = 0$. In fact, one can show that any $\alpha \in H^3(X, Z)$ of order dividing n arises in this way.

Can we define an appropriate K -theory for projective bundles with $\alpha \neq 0$? The answer is yes. For each fixed α of finite order we can define an Abelian group $K_\alpha(X)$. Moreover this is a $K(X)$ module. We will now indicate how these *twisted K -groups* (i.e., twisted K -theory) can be defined.

Note first that, for any vector space V , $\text{End } V = V \otimes V^*$ depends only $P(V)$. Hence, given a projective bundle P over X , we can define the associated bundle $\mathcal{E}(P)$ of endomorphism (matrix) algebras. The sections of $\mathcal{E}(P)$ form a non-commutative C^* -algebra and one can define its K -group by using finitely-generated projective modules. This K -group turns out to depend not on P but only on its obstruction class $\alpha \in H^3(X, Z)$ and so can be denoted by $K_\alpha(X)$.

In addition to the $K(X)$ –module structure of $K_\alpha(X)$ there are multiplications

$$K_\alpha(X) \otimes K_\beta(X) \rightarrow K_{\alpha+\beta}(X).$$

Today, there are many variants and generalizations of K –theory, something which is not surprising given the universality of linear algebra and matrices [Atiyah and Anderson (1967); Atiyah (2000)]. In each case there are specific features and techniques relevant to the particular area.

First, as already mentioned, is the *real* K –theory based on real vector bundles and the Bott periodicity theorems for the orthogonal groups: here the period is 8 rather than 2.

Next there is *equivariant theory* $K_G(X)$, where G is a compact Lie group acting on the space X . If X is a point, we just get the representation or character ring $R(G)$ of the group G . In general $K_G(X)$ is a module over $R(G)$ and this can be exploited in terms of the fixed–point sets in X of elements of G .

If we pass from the space X to the ring $C(X)$ of continuous complex–valued functions on X then $K(X)$ can be defined purely algebraically in terms of finitely–general projective modules over X . This then lends itself to a major generalization if we replace $C(X)$, which is a commutative C^* –algebra, by a *non–commutative* C^* –algebra. This has become a rich theory linked to many basic ideas in functional analysis, in particular to the *von Neumann dimension theory*.

4.5.5 Atiyah–Singer Index Theorem

We shall recall here very briefly some essential results of *Atiyah–Singer Index Theory*. The reader who is not familiar with the topological and analytic properties of the index of elliptic operators is urged to gain some familiarity with the *Atiyah–Singer Index Theorem* [Atiyah and Singer (1963); Atiyah and Singer (1968)]⁸ (for technical details, see also [Boos and Bleeker

⁸In the geometry of manifolds and differential operators, the Atiyah–Singer Index Theorem is an important unifying result that connects topology and analysis. It deals with elliptic differential operators (such as the Laplacian) on compact manifolds. It finds numerous applications, including many in theoretical physics. When Michael Atiyah and Isadore Singer were awarded the Abel Prize by the Norwegian Academy of Science and Letters in 2004, the prize announcement explained the Atiyah–Singer Index Theorem in these words:

“Scientists describe the world by measuring quantities and forces that vary over time and space. The rules of nature are often expressed by formulas, called

(1985)]).

A differential operator of order m , mapping the smooth sections of a vector bundle E over a compact manifold Y to those of another such bundle F , can be described in local coordinates and local trivializations of the bundles as

$$D = \sum_{|\alpha| \leq m} a_\alpha(x) D^\alpha,$$

with $\alpha = (\alpha_1, \dots, \alpha_n)$. The coefficients $a_\alpha(x)$ are matrices of smooth functions that represent elements of $\text{Hom}(E, F)$ locally; and $D^\alpha = \frac{\partial}{\partial x_1^{\alpha_1}} \cdots \frac{\partial}{\partial x_n^{\alpha_n}}$.

The principal symbol associated to the operator D is the expression

$$\sigma_m(D)(x, p) = \sum_{|\alpha|=m} a_\alpha(x) p^\alpha.$$

Given the differential operator $D : \Gamma(Y, E) \rightarrow \Gamma(Y, F)$, the principal symbol with the local expression above defines a global map $\sigma_m : \pi^*(E) \rightarrow \pi^*(F)$, where $T^*Y \xrightarrow{\pi} Y$ is the cotangent bundle; that is, the variables (x, p) are local coordinates on T^*Y .

Consider bundles E_i , $i = 1 \dots k$, over a compact n D manifold Y such that there is a complex $\Gamma(Y, E)$ formed by the spaces of sections $\Gamma(Y, E_i)$ and differential operators d_i of order m :

$$0 \rightarrow \Gamma(Y, E_1) \xrightarrow{d_1} \cdots \xrightarrow{d_{k-1}} \Gamma(Y, E_k) \rightarrow 0.$$

Construct the principal symbols $\sigma_m(d_i)$; these determine an associated symbol complex

$$0 \rightarrow \pi^*(E_1) \xrightarrow{\sigma_m(d_1)} \cdots \xrightarrow{\sigma_m(d_{k-1})} \pi^*(E_k) \rightarrow 0.$$

The complex $\Gamma(Y, E)$ is elliptic iff the associated symbol complex is exact.

differential equations, involving their rates of change. Such formulas may have an ‘index’, the number of solutions of the formulas minus the number of restrictions that they impose on the values of the quantities being computed. The Atiyah–Singer index Theorem calculated this number in terms of the geometry of the surrounding space. A simple case is illustrated by a famous paradoxical etching of M. C. Escher, ‘Ascending and Descending’, where the people, going uphill all the time, still manage to circle the castle courtyard. The index Theorem would have told them this was impossible.”

In the case of just one operator, this means that $\sigma_m(d)$ is an isomorphism off the zero section.

Recall that the *Hodge Theorem* states that the cohomology of the complex $\Gamma(Y, E)$ coincides with the harmonic forms, i.e.,

$$H^i(E) = \frac{\text{Ker}(d_i)}{\text{Im}(d_{i-1})} \cong \text{Ker}(\Delta_i), \quad \text{where} \quad \Delta_i = d_i^* d_i + d_{i-1} d_{i-1}^*.$$

Without loss of generality, by passing to the assembled complex

$$E^+ = E^1 \oplus E^3 \oplus \cdots, \quad E^- = E^2 \oplus E^4 \oplus \cdots,$$

we can always think of one elliptic operator

$$D : \Gamma(Y, E^+) \rightarrow \Gamma(Y, E^-), \quad D = \sum_i (d_{2i-1} + d_{2i}^*).$$

The Index Theorem states: Consider an elliptic complex over a compact, orientable, even dimensional manifold Y without boundary. The index of D , which is given by

$$\text{Ind}(D) = \dim[\text{Ker}(D)] - \dim[\text{Coker}(D)] = \sum_i (-1)^i \dim[\text{Ker } \Delta_i] = -\chi(E),$$

$\chi(E)$ being the Euler characteristic of the complex, can be expressed in terms of characteristic classes as:

$$\text{Ind}(D) = (-1)^{n/2} \left\langle \frac{\text{ch}(\sum_i (-1)^i [E_i])}{e(Y)} \text{td}(TY_{\mathbb{C}}), [Y] \right\rangle.$$

Here, ch is the Chern character, e is the Euler class of the tangent bundle of Y , $\text{td}(TY_{\mathbb{C}})$ is the Todd class of the complexified tangent bundle. The Atiyah–Singer Index Theorem, which computes the index of a family of elliptic differential operators, is naturally formulated in terms of K -theory and is an extension of the Riemann–Roch Theorem.

4.5.6 The Infinite-Order Case

Topological K -theory turned out to have a very natural link with the theory of operators in *quantum Hilbert space*. If \mathcal{H} is an infinite-dimensional complex Hilbert space and $\mathcal{B}(\mathcal{H})$ the space of bounded operators on \mathcal{H} with the uniform norm, then one defines the subspace $\mathcal{F}(\mathcal{H}) \subset \mathcal{B}(\mathcal{H})$ of *Fredholm operators* T , by the requirement that both $\text{Ker}(T)$ and $\text{Coker}(T)$ have

finite-dimensions. The Atiyah–Singer index is then defined by

$$\text{Ind}(T) = \dim[\text{Ker}(T)] - \dim[\text{Coker}(T)],$$

and it has the key property that it is continuous, and therefore constant on each connected component of $\mathcal{F}(\mathcal{H})$. Moreover, $\text{Ind} : \mathcal{F} \rightarrow \mathbb{Z}$ identifies the components of \mathcal{F} .

This has a generalization to any compact space X . To any continuous map $X \rightarrow \mathcal{F}$ (i.e., a continuous family of Fredholm operators parametrized by X) one can assign an index in $K(X)$. Moreover one gets in this way an isomorphism $\text{Ind} : [X, \mathcal{F}] \rightarrow K(X)$, where $[X, \mathcal{F}]$ denotes the set of homotopy classes of maps of X into \mathcal{F} . A notable example of a Fredholm operator is an elliptic differential operator on a compact manifold (these are turned into bounded operators by using appropriate *Sobolev norms*).

Now, in the quantum-physical situation, one meets infinite-order elements $\alpha \in H^3(X, \mathbb{Z})$ and the question arises of whether one can still define a ‘twisted’ group $K_\alpha(X)$. It turns out that it is possible to do this and one approach is being developed by [Atiyah and Segal (1971)].

Since an α of order n arises from an obstruction problem involving n D vector bundles, it is plausible that, for α of infinite order, we need to consider bundles of Hilbert spaces \mathcal{H} . But here we have to be careful not to confuse the ‘small’ unitary group

$$U(\infty) = \lim_{N \rightarrow \infty} U(N)$$

with the ‘large’ group $U(\mathcal{H})$ of all unitary operators in Hilbert space. The small unitary group has interesting homotopy groups given by Bott’s periodicity Theorem, but $U(\mathcal{H})$ is contractible, by *Kuiper’s Theorem*. This means that all Hilbert space bundles (with $U(\mathcal{H})$) as structure group) are trivial. This implies the following homotopy equivalences:

$$PU(\mathcal{H}) = U(\mathcal{H})/U(1) \sim CP_\infty = K(\mathbb{Z}, 2), \quad BPU(\mathcal{H}) \sim K(\mathbb{Z}, 3),$$

where B denotes here the classifying space and on the right we have the *Eilenberg–MacLane spaces*. It follows that $P(\mathcal{H})$ –bundles over X are classified completely by $H^3(X, \mathbb{Z})$. Thus, for each $\alpha \in H^3(X, \mathbb{Z})$, there is an essentially unique bundle P_α over X with fibre $P(\mathcal{H})$.

As in finite dimensions. $\mathcal{B}(\mathcal{H})$ depends only on $P(\mathcal{H})$, we can define a bundle \mathcal{B}_α of algebras over X . We now let $\mathcal{F}_\alpha \subset \mathcal{B}_\alpha$ be the corresponding bundle of Fredholm operators. Finally we define

$$K_\alpha(X) = \text{Homotopy classes of sections of } \mathcal{F}_\alpha.$$

This definition works for all α . If α is of finite order, then P_α contains a finite-dimensional sub-bundle, but if α is of infinite order this is not true. Thus we are essentially in an infinite-dimensional analytic situation.

To get the twisted odd groups we recall that $\mathcal{F}^1 \subset \mathcal{F}$, the space of self-adjoint Fredholm operators, is a classifying space for K^1 and so we take $\mathcal{F}_\alpha^1 \subset \mathcal{F}_\alpha$ to define

$$K_\alpha^1(X) = \text{Homotopy classes of sections of } \mathcal{F}_\alpha^1.$$

One peculiar feature of the infinite-order case is that all sections of \mathcal{F}_α lie in the zero-index component, or equivalently that the restriction map

$$K_\alpha(X) \rightarrow K_\alpha(\text{point})$$

is zero.

Now, what can we say about the relation between twisted K -groups and cohomology? Over the rationals, if α is of finite order, nothing much changes [Atiyah and Anderson (1967); Atiyah (2000)]. In particular the *Chern character* induces an isomorphism. However if α is of infinite order something new happens. We can now consider the operation $u \rightarrow \alpha u$ on $H^*(X, Q)$ as a differential d_α ($\alpha^2 = 0$ since α has odd dimension). We can then form the cohomology with respect to this differential

$$\mathcal{H}_\alpha = \text{Ker } d_\alpha / \text{Im } d_\alpha.$$

One can then prove that there is an isomorphism

$$K_\alpha^*(X) \otimes Q \cong \mathcal{H}_\alpha.$$

In the usual Atiyah–Hirzebruch spectral sequence [Atiyah and Hirzebruch (1961)], relating K -theory to integral cohomology, all differentials are of finite order and so vanish over Q . In particular, $d_3 = Sq^3$, the *Steenrod operation*. However for K_α one finds

$$d_3 u = Sq^3 u + \alpha u$$

and this explains why an α of infinite-order gives the isomorphism above over the rationals.

Chern classes over the integers are a more delicate matter. One can proceed as follows. In \mathcal{F} there are various subspaces $\mathcal{F}_{r,s}$ (of finite codimension) where

$$\dim[\text{Ker}] = r, \quad \dim[\text{Coker}] = s,$$

and these lie in the component of $\text{Ind}(r - s)$.

Since the $\mathcal{F}_{r,s} \subset \mathcal{F}$ are invariant under the action of $U(\mathcal{H})$, it follows that they can be defined fibrewise and this shows that the classes $c_{r,s}$ can be defined for $K_\alpha(X)$. However the classes for $r = s$ (and so of index zero) are not sufficient to generate all Chern classes. It is a not unreasonable conjecture that the $c_{r,r}$ are the only integral characteristic classes for the twisted K -theories [Atiyah and Anderson (1967); Atiyah (2000)].

While the use of Hilbert spaces H and the corresponding projective spaces $P(\mathcal{H})$ may not come naturally to a topologist, these are perfectly natural in physics. Recall that $P(\mathcal{H})$ is the *space of quantum states*. Bundles of such arise naturally in quantum field theory.

4.5.7 Twisted K -Theory and the Verlinde Algebra

Twistings of cohomology theories are most familiar for ordinary cohomology [Freed (2001); Freed *et. al.* (2003)]. Let M be a smooth manifold. Then a flat real vector bundle $E \rightarrow M$ determines twisted real cohomology groups $H^\bullet(M; E)$. In differential geometry these cohomology groups are defined by extending the de Rham complex to forms with coefficients in E using the flat connection. The sorts of twistings of K -theory we consider are 1D, so analogous to the case when E is a line bundle. There are also 1D twistings of *integral* cohomology, determined by a *local system* $Z \rightarrow M$. This is a bundle of groups isomorphic to \mathbb{Z} , so is determined up to isomorphism by an element of $H^1(M; \text{Aut}(\mathbb{Z})) \cong H^1(M; \mathbb{Z} \bmod 2)$, since the only nontrivial automorphism of \mathbb{Z} is multiplication by -1 . The twisted integral cohomology $H^\bullet(M; Z)$ may be thought of as sheaf cohomology, or defined using a cochain complex. We give a Čech description as follows. Let $\{U_i\}$ be an open covering of M and

$$g_{ij} : U_i \cap U_j \longrightarrow \{\pm 1\} \quad (4.28)$$

a cocycle defining the local system Z . Then an element of $H^q(M; Z)$ is represented by a collection of q -cochains $a_i \in Z^q(U_i)$ which satisfy

$$a_j = g_{ij} a_i \quad \text{on} \quad U_{ij} = U_i \cap U_j. \quad (4.29)$$

We can use any model of co-chains, since the group $\text{Aut}(\mathbb{Z}) \cong \{\pm 1\}$ always acts. In place of co-chains we represent integral cohomology classes by maps to an Eilenberg–MacLane space $K(\mathbb{Z}, q)$. The cohomology group is the set of homotopy classes of maps, but here we use honest maps as representatives. The group $\text{Aut}(\mathbb{Z})$ acts on $K(\mathbb{Z}, q)$. One model of $K(\mathbb{Z}, 0)$ is the

integers, with -1 acting by multiplication. The circle is a model for $K(\mathbb{Z}, 1)$, and -1 acts by reflection. Using the action of $\text{Aut}(\mathbb{Z})$ on $K(\mathbb{Z}, q)$ and the cocycle (4.28) we build an associated bundle $\mathcal{H}^q \rightarrow M$ with fiber $K(\mathbb{Z}, q)$. Equation (4.29) says that twisted cohomology classes are represented by sections of $\mathcal{H}^q \rightarrow M$; the twisted cohomology group $H^q(M; \mathbb{Z})$ is the set of homotopy classes of sections of $\mathcal{H}^q \rightarrow M$.

Twistings may be defined for any generalized cohomology theory; our interest is in complex K -theory [Freed (2001); Freed *et. al.* (2003)]. In homotopy theory one regards K as a marriage of a ring and a space (more precisely, spectrum), and it makes sense to ask for the units in K , denoted $GL_1(K)$. In the previous paragraph we used the units in integral cohomology, the group $\mathbb{Z} \bmod 2$. For complex K -theory there is a richer group of units

$$GL_1(K) \sim \mathbb{Z} \bmod 2 \times CP^\infty \times BSU. \quad (4.30)$$

In our problem the last factor does not enter and all the interest is in the first two, which we denote $GL_1(K)'$. As a first approximation, view K as the category of all finite dimensional $\mathbb{Z} \bmod 2$ -graded complex vector spaces. Then CP^∞ is the subcategory of even complex lines, and it is a group under tensor product. It acts on K by tensor product as well. The nontrivial element of $\mathbb{Z} \bmod 2$ in (4.30) acts on K by reversing the parity of the grading. This model is deficient since there is not an appropriate topology. One may consider instead complexes of complex vector spaces, or spaces of operators as we do below. Of course, there are good topological models of CP^∞ , for example the space of all 1D subspaces of a fixed complex Hilbert space \mathcal{H} . For a manifold M the twistings of K -theory of interest are classified up to isomorphism by

$$H^1(M; GL_1(K)') \cong H^1(M; \mathbb{Z} \bmod 2) \times H^3(M; \mathbb{Z}).$$

In this paper we will not encounter twistings from the first factor and will focus exclusively on the second. These twistings are represented by cocycles g_{ij} with values in the space of lines, in other words by complex line bundles $L_{ij} \rightarrow U_{ij}$ which satisfy a cocycle condition. This is the data often given to define a *gerbe*.⁹

⁹Recall that a *gerbe* is a construct in homological algebra. It is defined as a *stack* over a topological space which is locally isomorphic to the *Picard groupoid* of that space. Recall that the Picard groupoid on an open set U is the category whose objects are *line bundles* on U and whose morphisms are isomorphisms. A stack refers to any category acting like a *moduli space* with a universal family (analogous to a classifying space)

Now, let $X = G$ be a compact Lie group and, for simplicity, we shall assume that it is simply connected, though the theory works in the general case. We consider G as G -space, the group acting on itself by conjugation.

Since $H^3(G, \mathbb{Z}) \cong \mathbb{Z}$ we can construct twisted K -theories for each integer k . Moreover, we can also do this equivariantly, thus obtaining Abelian groups $K_{G,k}^*(G)$. These will all be $R(G)$ -modules.

Now, the group multiplication map $\mu : G \times G \rightarrow G$ is compatible with conjugation and so is a G -map. In addition, to the pull back μ^* , we can also consider the push-forward μ_* . This depends on Poincaré duality for K -theory and it works also, when appropriately formulated, in the present context.

If $\dim(G)$ is even, this gives us a commutative multiplication on $K_{G,k}^0(G)$, while for $\dim(G)$ odd, our multiplication is on $K_{G,k}^1(G)$. In either case we get a ring.

The claim of [Freed (2001); Freed *et. al.* (2003)] is that this ring (according to the parity of $\dim(G)$ is naturally isomorphic to the *Verlinde algebra* of G at level $k - h$ (where h is the Coxeter number). The Verlinde algebra is a key tool in certain quantum field theories and it has been much studied by physicists, topologists, group theorists and algebraic geometers. The K -theory approach is totally new and much more direct than most other ways.

The Verlinde algebra is defined in the theory of loop groups. Let G be a compact Lie group. There is a version of the Theorem for *any* compact group G , but here for the most part we focus on connected, simply connected, and simple groups— $G = SU_2$ is the simplest example. In this case a central extension of the free loop group LG is determined by the *level*, which is a positive integer k . There is a finite set of equivalence classes of positive energy representations of this central extension; let $V_k(G)$ denote the free Abelian group they generate. One of the influences of 2D conformal field theory on the theory of loop groups is the construction of an algebra structure on $V_k(G)$, the *fusion product*. This is the *Verlinde algebra* [Verlinde (1988)].

More precisely, let G act on itself by conjugation. Then with our assumptions the equivariant cohomology group $H_G^3(G)$ is free of rank one. Let $h(G)$ be the dual *Coxeter number* of G , and define $\zeta(k) \in H_G^3(G)$ to be $k + h(G)$ times a generator. We will see that elements of H^3 may be used to *twist* K -theory, and so elements of equivariant H^3 twist equivariant

parameterizing a family of related mathematical objects such as schemes or topological spaces, especially when the members of these families have nontrivial automorphisms.

K -theory.

The *Freed–Hopkins–Teleman Theorem* [Freed (2001); Freed *et. al.* (2003)] states: There is an isomorphism of algebras

$$V_k(G) \cong K_G^{\dim G + \zeta(k)}(G),$$

where the right hand side is the $\zeta(k)$ -twisted equivariant K -theory in degree $\dim(G)$. The group structure on the right-hand side is induced from the multiplication map $G \times G \rightarrow G$.

For an arbitrary compact Lie group G the level k is replaced by a class in $H^4(BG; \mathbb{Z})$ and the dual Coxeter number $h(G)$ is pulled back from a universal class in $H^4(BSO; \mathbb{Z})$ via the adjoint representation. The twisting class is obtained from their sum by transgression.

4.5.8 Application: K -Theory in String Theory

In string theory, K -theory has been conjectured to classify the allowed *Ramond–Ramond field* strengths;¹⁰ and also the charges of stable D -branes.

4.5.8.1 Classification of Ramond–Ramond Fluxes

In the classical limit of type II string theory, which is type II supergravity, the Ramond–Ramond (RR) field strengths are differential forms. In the quantum theory the well-definedness of the partition functions of D -branes implies that the RR-field strengths obey *Dirac quantization* conditions when space–time is compact, or when a spatial slice is compact and one considers only the (magnetic) components of the field strength which lie along the spatial directions. This led twentieth century physicists to classify RR field strengths using cohomology with integral coefficients.

However, some authors have argued that the cohomology of space–time with integral coefficients is too big. For example, in the presence of Neveu–Schwarz (NS) H -flux, or non-spin cycles, some RR-fluxes dictate the presence of D -branes. In the former case this is a consequence of the super-

¹⁰Recall that Ramond–Ramond (RR) fields are differential-form fields in the 10D space–time of type II supergravity theories, which are the classical limits of type II string theory. The ranks of the fields depend on which type II theory is considered. As Joe Polchinski argued in 1995, D -branes are the charged objects that act as sources for these fields, according to the rules of p -form electrodynamics. It has been conjectured that quantum RR fields are not differential forms, but instead are classified by twisted K -theory.

gravity equation of motion which states that the product of a RR-flux with the NS 3-form is a D -brane charge density. Thus the set of topologically distinct RR-field strengths that can exist in brane-free configurations is only a subset of the cohomology with integral coefficients.

This subset is still too big, because some of these classes are related by large gauge transformations. In QED there are large gauge transformations which add integral multiples of 2π to *Wilson loops*.¹¹

¹¹Recall that in gauge theory, a *Wilson loop* (named after Ken Wilson) is a gauge-invariant observable obtained from the holonomy of the gauge connection around a given loop. In the classical theory, the collection of all Wilson loops contains sufficient information to reconstruct the gauge connection, up to gauge transformation. In *quantum field theory*, the definition of Wilson loops observables as *bona fide* operators on *Fock space* (actually, Haag's Theorem states that Fock space does not exist for interacting QFTs) is a mathematically delicate problem and requires regularization, usually by equipping each loop with a framing. The action of Wilson loop operators has the interpretation of creating an elementary excitation of the quantum field which is localized on the loop. In this way, Faraday's "flux tubes" become elementary excitations of the quantum electromagnetic field.

Wilson loops were introduced in the 1970s in an attempt at a non-perturbative formulation of quantum chromodynamics (QCD), or at least as a convenient collection of variables for dealing with the strongly-interacting regime of QCD. The problem of confinement, which Wilson loops were designed to solve, remains unsolved to this day. The fact that strongly-coupled quantum gauge field theories have elementary non-perturbative excitations which are loops motivated Alexander Polyakov to formulate the first string theories, which described the propagation of an elementary quantum loop in spacetime.

Wilson loops played an important role in the formulation of *loop quantum gravity*, but there they are superseded by spin networks, a certain generalization of Wilson loops.

In particle physics and string theory, Wilson loops are often called *Wilson lines*, especially Wilson loops around non-contractible loops of a compact manifold.

A Wilson line W_C is a quantity defined by a *path-ordered exponential* of a *gauge field* A_μ

$$W_C = \text{Tr } \mathcal{P} \exp i \oint_C A_\mu dx^\mu.$$

Here, C is a contour in space, \mathcal{P} is the path-ordering operator, and the trace Tr guarantees that the operator is invariant under gauge transformations. Note that the quantity being traced over is an element of the *gauge Lie group* and the trace is really the character of this element with respect to an *irreducible representation*, which means there are infinitely many traces, one for each *irrep*.

Precisely because we're looking at the trace, it doesn't matter which point on the loop is chosen as the initial point. They all give the same value.

Actually, if A is viewed as a connection over a principal G -bundle, the equation above really ought to be 'read' as the *parallel transport* of the identity around the loop which would give an element of the Lie group G .

Note that a path-ordered exponential is a convenient shorthand notation common in physics which conceals a fair number of mathematical operations. A mathematician would refer to the path-ordered exponential of the connection as 'the holonomy of the connection' and characterize it by the parallel-transport differential equation that it satisfies.

The p -form potentials in type II supergravity theories also enjoy these large gauge transformations, but due to the presence of Chern–Simons terms in the supergravity actions these large gauge transformations transform not only the p -form potentials but also simultaneously the $(p+3)$ -form field strengths. Thus to get the space of inequivalent field strengths from the forementioned subset of integral cohomology we must quotient by these large gauge transformations.

The Atiyah–Hirzebruch spectral sequence constructs twisted K -theory, with a twist given by the NS 3-form field strength, as a quotient of a subset of the cohomology with integral coefficients. In the classical limit, which corresponds to working with rational coefficients, this is precisely the quotient of a subset described above in supergravity. The quantum corrections come from torsion classes and contain mod 2 torsion corrections due to the Freed–Witten anomaly.

Thus twisted K -theory classifies the subset of RR-field strengths that can exist in the absence of D -branes quotiented by large gauge transformations.

4.5.8.2 Classification of D -Branes

Now, in many applications one wishes to add sources for the RR fields. These sources are called D -branes. As in classical electromagnetism, one may add sources by including a coupling $Cp\mathcal{J}_{10-p}$ of the p -form *potential* to a $(10-p)$ -form *current* \mathcal{J}_{10-p} in the Lagrangian (density). The usual convention in the string theory literature appears to be to not write this term explicitly in the action.

The current \mathcal{J}_{10-p} modifies the equation of motion that comes from the variation of Cp . As is the case with magnetic monopoles in electromagnetism, this source also invalidates the dual *Bianchi identity* as it is a point at which the dual field is not defined. In the modified equation of motion \mathcal{J}_{p+2} appears on the left hand side of the equation of motion instead of zero. For simplicity, we will also interchange p and $7-p$, then the equation of motion in the presence of a source is

$$\mathcal{J}_{9-p} = d^2 C_{7-p} = dG_{8-p} = dF_{8-p} + H \wedge G_{p-1}.$$

The $(9-p)$ -form \mathcal{J}_{9-p} is the Dp -brane current, which means that it is *Poincaré dual* to the world-volume of a $(p+1)$ -D extended object called

In finite temperature QCD, the expectation value of the Wilson line distinguishes between the ‘confined phase’ and the ‘deconfined phase’ of the theory.

a Dp -brane. The discrepancy of one in the naming scheme is historical and comes from the fact that one of the $p + 1$ directions spanned by the Dp -brane is often time-like, leaving p spatial directions.

The above Bianchi identity is interpreted to mean that the Dp -brane is, in analogy with magnetic monopoles in electromagnetism, magnetically charged under the RR p -form $C7 - p$. If instead one considers this Bianchi identity to be a field equation for $Cp + 1$, then one says that the Dp -brane is electrically charged under the $(p + 1)$ -form $Cp + 1$.

The above equation of motion implies that there are two ways to derive the Dp -brane charge from the ambient fluxes. First, one may integrate dG_{8-p} over a surface, which will give the Dp -brane charge intersected by that surface. The second method is related to the first by *Stoke's Theorem*. One may integrate G_{8-p} over a cycle, this will yield the Dp -brane charge linked by that cycle. The quantization of Dp -brane charge in the quantum theory then implies the quantization of the field strengths G , but not of the improved field strengths F .

It has been conjectured that twisted K -theory classifies D -branes in noncompact space-times, intuitively in space-times in which we are not concerned about the flux sourced by the brane having nowhere to go. While the K -theory of a 10D space-time classifies D -branes as subsets of that space-time, if the space-time is the product of time and a fixed 9-manifold then K -theory also classifies the conserved D -brane charges on each 9D spatial slice. While we were required to forget about RR potentials to get the K -theory classification of RR field strengths, we are required to forget about RR field strengths to get the K -theory classification of D -branes.

We will continue exposition on K -theory applications to string theory in the last section of the book.

4.6 Principal Bundles

Recall that a *principal bundle* is a special case of a fibre bundle where the fibre is a group G . More specifically, G is usually a Lie group. A principal bundle is a total bundle space Y along with a surjective map $\pi : Y \rightarrow X$ to a base manifold X . Any fibre $\pi^{-1}(x)$ is a space isomorphic to G . More specifically, G acts freely and transitively without fixed point on the fibers, and this makes a fibre into a homogeneous space. It follows that the orbits of the G -action are precisely the fibers of $\pi : Y \rightarrow X$ and the orbit space

Y/G is homeomorphic the base space X . To say that G acts freely and transitively on the fibers means that the fibers take on the structure of G -torsors.¹²

For example, in the case of a *circle bundle* ($G = S^1 \equiv \{e^{it}\}$), the fibers are circles, which can be rotated, although no point in particular corresponds to the identity. Near every point, the fibers can be given the group structure of G in the fibers over a neighborhood by choosing an element in each fibre to be the identity element. However, the fibers cannot be given a group structure globally, except in the case of a trivial bundle.

An important principal bundle is the *frame bundle* on a Riemannian manifold. This bundle reflects the different ways to give an orthonormal basis for tangent vectors.

In general, any fibre bundle corresponds to a principal bundle where the group (of the principal bundle) is the group of isomorphisms of the fibre (of the fibre bundle). Given a principal bundle $\pi : Y \rightarrow X$ and an action of G on a space V , which could be a group representation, this can be reversed to give an *associated* fibre bundle.

A trivialization of a principal bundle, an open set U in X such that the bundle $\pi^{-1}(U)$ over U , is expressed as $U \times G$, has the property that the group G acts on the left and transition functions take values in G , acting on the fibers by right multiplication (so that the action of G on a fibre V is independent of coordinate chart).

More precisely, a principal bundle $\pi_P : P \rightarrow Q$ of a configuration manifold Q , with a *structure Lie group* G , is a general affine bundle modelled on the right on the trivial group bundle $Q \times G$ where the group G acts freely and transitively on P on the right,

$$R_G : P \times G \rightarrow P, \quad R_g : p \mapsto pg, \quad (p \in P, g \in G). \quad (4.31)$$

We call P a *principal G -bundle*. A typical fibre of a principal G -bundle is isomorphic to the group space of G , and $P/G = Q$. The structure group G acts on the typical fibre by left multiplications which do not preserve the group structure of G . Therefore, the typical fibre of a principal bundle is only a group space, but not a group. Since the left action of transition functions on the typical fibre G commutes with its right multiplications, a principal bundle admits the global right action (4.31) of the structure group.

¹²A G -torsor is a space which is homeomorphic to G but lacks a group structure since there is no preferred choice of an identity element.

A principal G -bundle P is equipped with a bundle atlas

$$\Psi_P = \{(U_\alpha, \psi_\alpha^P, \rho_{\alpha\beta})\}, \quad (4.32)$$

whose trivialization maps

$$\psi_\alpha^P : \pi_P^{-1}(U_\alpha) \rightarrow U_\alpha \times G$$

obey the *equivariance condition*

$$(\pi_2 \circ \psi_\alpha^P)(pg) = (\pi_2 \circ \psi_\alpha^P)(p)g, \quad (g \in G, p \in \pi_P^{-1}(U_\alpha)). \quad (4.33)$$

Due to this property, every trivialization map ψ_α^P determines a unique local section z_α of P over U_α such that

$$\pi_2 \circ \psi_\alpha^P \circ z_\alpha = \mathbf{1},$$

where $\mathbf{1}$ is the unit element of G . The transformation rules for z_α read

$$z_\beta(q) = z_\alpha(q)\rho_{\alpha\beta}(q), \quad (q \in U_\alpha \cap U_\beta), \quad (4.34)$$

where $\rho_{\alpha\beta}(q)$ are G -valued transition functions of the atlas Ψ_P . Conversely, the family $\{(U_\alpha, z_\alpha)\}$ of local sections of P with the transition functions (4.34) determines a unique bundle atlas of P . In particular, it follows that only trivial principal bundles have global sections.

Note that the pull-back of a principal bundle is also a principal bundle with the same structure group.

The quotient of the tangent bundle $TP \rightarrow P$ and that of the vertical tangent bundle VP of P by the tangent prolongation TR_G of the canonical action R_G (4.31) are vector bundles

$$T_GP = TP/G, \quad V_GP = VP/G \quad (4.35)$$

over Q . Sections of $T_GP \rightarrow Q$ are naturally identified with G -invariant vector-fields on P , while those of $V_GP \rightarrow Q$ are G -invariant vertical vector-fields on P . Therefore, the Lie bracket of G -invariant vector-fields on P goes to the quotients (4.35), and induces the Lie brackets of their sections. Let us write these brackets in an explicit form.

Owing to the equivariance condition (4.33), any bundle atlas (4.32) of P induces the associated bundle atlases $\{U_\alpha, T\psi_\alpha^P/G\}$ of T_GP and $\{U_\alpha, V\psi_\alpha^P/G\}$ of V_GP . Given a basis $\{\varepsilon_p\}$ for the right Lie algebra \mathfrak{g}_r , let $\{\partial_\alpha, e_p\}$ and $\{e_p\}$, where $e_p = (\psi_\alpha^P/G)^{-1}(\varepsilon_p)$, be the corresponding local fibre bases for the vector bundles T_GP and V_GP , respectively. Relative

to these bases, the Lie bracket of sections

$$\xi = \xi^\alpha \partial_\alpha + \xi^p e_p, \quad \eta = \eta^\mu \partial_\mu + \eta^q e_q$$

of the vector bundle $T_G P \rightarrow Q$ reads

$$[\xi, \eta] = (\xi^\mu \partial_\mu \eta^\alpha - \eta^\mu \partial_\mu \xi^\alpha) \partial_\alpha + (\xi^\alpha \partial_\alpha \eta^r - \eta^\alpha \partial_\alpha \xi^r + c_{pq}^r \xi^p \eta^q) e_r. \quad (4.36)$$

Putting $\xi^\alpha = 0$ and $\eta^\mu = 0$, we get the Lie bracket

$$[\xi, \eta] = c_{pq}^r \xi^p \eta^q e_r \quad (4.37)$$

of sections of the vector bundle $V_G \rightarrow P$.

A principal bundle P is also the general affine bundle modelled on the left on the associated group bundle \tilde{P} with the standard fibre G on which the structure group G acts by the adjoint representation. The corresponding bundle map reads

$$\tilde{P} \times P \rightarrow P, \quad (\tilde{p}, p) \mapsto \tilde{p}p.$$

Note that the standard fibre of the group bundle \tilde{P} is the group G , while that of the principal bundle P is the group space of G on which the structure group G acts on the left.

A principal bundle $P \rightarrow Q$ with a structure Lie group G possesses the canonical trivial vertical splitting

$$\alpha : VP \rightarrow P \times \mathfrak{g}_l, \quad \pi_2 \circ \alpha \circ e_m = J_m,$$

where $\{J_m\}$ is a basis for the left Lie algebra \mathfrak{g}_l and e_m denotes the corresponding fundamental vector-fields on P . Given a principal bundle $P \rightarrow Q$, the bundle $TP \rightarrow TQ$ is a principal bundle

$$TP \times T(Q \times G) \rightarrow TP$$

with the structure group $TG = G \times \mathfrak{g}_l$ where \mathfrak{g}_l is the left Lie algebra of left-invariant vector-fields on the group G .

If $P \rightarrow Q$ is a principal bundle with a structure group G , the exact sequence (4.13) can be reduced to the exact sequence

$$0 \rightarrow V^G P \hookrightarrow T^G P \rightarrow TQ \rightarrow 0, \quad (4.38)$$

where $T^G P = TP/G, \quad V^G P = VP/G$

are the quotients of the tangent bundle TP of P and the vertical tangent bundle VP of P respectively by the canonical action (4.31) of G on P on the right. The bundle $V^G P \rightarrow Q$ is called the adjoint bundle. Its standard fibre

is the right Lie algebra \mathfrak{g}_r of the right-invariant vector-fields on the group G . The group G acts on this standard fibre by the adjoint representation.

4.7 Distributions and Foliations on Manifolds

Let M be an n D smooth manifold. A *smooth distribution* \mathcal{T} of codimension k on M is defined as a subbundle of rank $n - k$ of the tangent bundle TM . A smooth distribution \mathcal{T} is called the *involutive distribution* if $[u, u']$ is a section of \mathcal{T} whenever u and u' are sections of \mathcal{T} .

Let \mathcal{T} be a k -codimensional distribution on M . Its annihilator \mathcal{T}^* is a k D subbundle of T^*M called the *Pfaffian system*. It means that, on a neighborhood U of every point $x \in M$, there exist k linearly independent sections s_1, \dots, s_k of \mathcal{T}^* such that

$$\mathcal{T}_x|_U = \cap_j \text{Ker } s_j.$$

Let $\mathcal{C}(\mathcal{T})$ be the ideal of $\wedge(M)$ generated by sections of \mathcal{T}^* .

A smooth distribution \mathcal{T} is involutive iff the *ideal* $\mathcal{C}(\mathcal{T})$ is *differential*, that is, $d\mathcal{C}(\mathcal{T}) \subset \mathcal{C}(\mathcal{T})$.

Given an involutive k -codimensional distribution \mathcal{T} on M , the quotient TM/\mathcal{T} is a k D vector bundle called the *transversal bundle* of \mathcal{T} . There is the exact sequence

$$0 \rightarrow \mathcal{T} \hookrightarrow TM \rightarrow TM/\mathcal{T} \rightarrow 0. \quad (4.39)$$

Given a bundle $Y \rightarrow X$, its vertical tangent bundle VY exemplifies an involutive distribution on Y .

A submanifold N of M is called the *integral manifold* of a distribution \mathcal{T} on M if the tangent spaces to N coincide with the fibres of this distribution at each point of N .

Let \mathcal{T} be a smooth involutive distribution on M . For any point $x \in M$, there exists a maximal integral manifold of \mathcal{T} passing through x [Kamber and Tondeur (1975)]. In view of this fact, involutive distributions are also called *completely integrable distributions*.

Every point $x \in M$ has an open neighborhood U which is a domain of a coordinate chart (x^1, \dots, x^n) such that the restrictions of \mathcal{T} and \mathcal{T}^* to U are generated by the $n - k$ vector-fields $\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^{n-k}}$ and the k Pfaffian forms dx^{n-k+1}, \dots, dx^n respectively.

In particular, it follows that integral manifolds of an involutive distribution constitute a *foliation*. Recall that a k -codimensional foliation on

an n D manifold M is a partition of M into *connected leaves* F_l with the following property: every point of M has an open neighborhood U which is a domain of a coordinate chart (x^α) such that, for every leaf F_l , the components $F_l \cap U$ are described by the equations $x^{n-k+1} = \text{const}, \dots, x^n = \text{const}$ [Kamber and Tondeur (1975)]. Note that leaves of a foliation fail to be imbedded submanifolds in general.

For example, every projection $\pi : M \rightarrow X$ defines a foliation whose leaves are the fibres $\pi^{-1}(x)$, for all $x \in X$. Also, every nowhere vanishing vector-field u on a manifold M defines a 1D involutive distribution on M . Its integral manifolds are the integral curves of u . Around each point $x \in M$, there exist local coordinates (x^1, \dots, x^n) of a neighborhood of x such that u is given by $u = \frac{\partial}{\partial x^1}$.

4.8 Application: Nonholonomic Mechanics

Let $TM = \cup_{x \in M} T_x M$, be the tangent bundle of a smooth n D mechanical manifold M . Recall (from the subsection 4.7 above) that sub-bundle $V = \cup_{x \in M} V_x$, where V_x is a vector subspace of $T_x M$, smoothly dependent on points $x \in M$, is called the *distribution*. If the manifold M is connected, $\dim V_x$ is called the dimension of the distribution. A vector-field X on M belongs to the distribution V if $X(x) \subset V_x$. A curve γ is *admissible relatively to* V , if the vector-field $\dot{\gamma}$ belongs to V . A *differential system* is a linear space of vector-fields having a structure of $C^\infty(M)$ - module. Vector-fields which belong to the distribution V form a differential system $N(V)$. A k D distribution V is *integrable* if the manifold M is foliated to k D sub-manifolds, having V_x as the tangent space at the point x . According to the Frobenius Theorem, V is integrable iff the corresponding differential system $N(V)$ is *involutive*, i.e., if it is a Lie sub-algebra of the Lie algebra of vector-fields on M . The *flag* of a differential system N is a sequence of differential systems: $N_0 = N$, $N_1 = [N, N], \dots, N_l = [N_{l-1}, N], \dots$.

The differential systems N_i are not always differential systems of some distributions V_i , but if for every i , there exists V_i , such that $N_i = N(V_i)$, then there exists a *flag* of the distribution V : $V = V_0 \subset V_1 \dots$. Such distributions, which have flags, will be called *regular*. Note that the sequence $N(V_i)$ is going to stabilize, and there exists a number r such that $N(V_{r-1}) \subset N(V_r) = N(V_{r+1})$. If there exists a number r such that $V_r = TM$, the distribution V is called *completely nonholonomic*, and minimal such r is the *degree of non-holonomicity* of the distribution V .

Now, let us see the mechanical interpretation of these geometrical objects. Consider a nonholonomic mechanical system corresponding to a Riemannian manifold (M, g) , where g is a metric defined by the system's kinetic energy [Dragovic and Gajic (2003)]. Suppose that the distribution V is defined by $(n - m)$ one-forms ω_α ; in local coordinates $q = (q^1, \dots, q^n)$ on M

$$\omega_\rho(q)(\dot{q}) = a_{\rho i}(q) \dot{q}^i = 0, \quad (\rho = m + 1, \dots, n; i = 1, \dots, n).$$

A *virtual displacement* is a vector-field X on M , such that $\omega_\rho(X) = 0$, i.e., X belongs to the differential system $N(V)$.

Differential equations of motion of a given mechanical system follow from the *D'Alembert-Lagrangian principle*: trajectory γ of the given system is a solution of the equation

$$\langle \nabla_{\dot{\gamma}} \dot{\gamma} - Q, X \rangle = 0, \quad (4.40)$$

where X is an arbitrary virtual displacement, Q a vector-field of internal forces, and ∇ is the affine *Levi-Civita connection* for the metric g .

The vector-field $R(x)$ on M , such that $R(x) \in V_x^\perp$, $V_x^\perp \oplus V_x = T_x M$, is called *reaction of ideal nonholonomic connections*. (4.40) can be rewritten as

$$\nabla_{\dot{\gamma}} \dot{\gamma} - Q = R, \quad \omega_\alpha(\dot{\gamma}) = 0. \quad (4.41)$$

If the system is potential, by introducing $L = T - U$, where U is the potential energy of the system ($Q = -\text{grad } U$), then in local coordinates q on M , equations (4.41) becomes the *forced Lagrangian equation*:

$$\frac{d}{dt} L_{\dot{q}} - L_q = \tilde{R}, \quad \omega_\alpha(\dot{q}) = 0.$$

Now \tilde{R} is a one-form in (V^\perp) , and it can be represented as a linear combination of one-forms $\omega^{m+1}, \dots, \omega^n$ which define the distribution, $\tilde{R} = \lambda_\alpha \omega_\alpha$.

Suppose e_1, \dots, e_n are the vector-fields on M , such that $e_1(x), \dots, e_n(x)$ form a base of the vector space $T_x M$ at every point $x \in M$, and e_1, \dots, e_m generate the differential system $N(V)$. Express them through the coordinate vector-fields:

$$e_i = A_i^j(q) \partial_{q^j}, \quad (i, j = 1, \dots, n).$$

Denote by p a projection $p : TM \rightarrow V$ orthogonal according to the metric g . Corresponding homomorphism of C^∞ -modules of sections of

TM and V is

$$p\tilde{\partial}_{q^i} = p_i^a e_a, \quad (a = 1, \dots, m, i = 1, \dots, n).$$

Projecting by p the equations (4.41), from $R(x) \in V^\perp(x)$, we get $p(R) = 0$, and denoting $p(Q) = \tilde{Q}$ we get

$$\tilde{\nabla}_{\dot{\gamma}} \dot{\gamma} = \tilde{Q},$$

where $\tilde{\nabla}$ is the *projected connection* [Dragovic and Gajic (2003)]. A relationship between standard Christoffel symbols Γ_{ij}^k and coefficients $\tilde{\Gamma}_{ab}^c$ of the connection $\tilde{\nabla}$, defined by

$$\tilde{\nabla}_{e_a} e_b = \tilde{\Gamma}_{ab}^c e_c, \quad \text{is given by} \quad \tilde{\Gamma}_{ab}^c = \Gamma_{ij}^k A_a^i A_b^j p_k^c + A_a^i \partial_{q^i} A_b^j p_j^c.$$

If the motion takes place under the inertia ($Q = \tilde{Q} = 0$), the trajectories of nonholonomic mechanical problem are the geodesics for $\tilde{\nabla}$.

Now, let V be a distribution on M . Denote a $C^\infty(M)$ -module of sections on V by $\Gamma(V)$. A *nonholonomic connection* on the sub-bundle V of TM is a map $\nabla : \Gamma(V) \times \Gamma(V) \rightarrow \Gamma(V)$ with the properties:

$$\begin{aligned} \nabla_X(Y + Z) &= \nabla_X Y + \nabla_X Z, & \nabla_X(f \cdot Y) &= X(f)Y + f\nabla_X Y, \\ \nabla_{fX+gY} Z &= f\nabla_X Z + g\nabla_Y Z, & (X, Y, Z \in \Gamma(V); f, g \in C^\infty(M)). \end{aligned}$$

Having a morphism of vector bundles $p_0 : TM \rightarrow V$, formed by the projection on V , denote by $q_0 = 1_{TM} - p_0$ the projection on W , $V \oplus W = TM$.

The tensor-field $T_\nabla : \Gamma(V) \times \Gamma(V) \rightarrow \Gamma(V)$ defined by

$$T_\nabla(X, Y) = \nabla_X Y - \nabla_Y X - p_0[X, Y], \quad (X, Y \in \Gamma(V)),$$

is called the *torsion tensor* for the connection ∇ .

Suppose there is a positively defined metric tensor $g = g_{ij}$ on V . Given a distribution V , with p_0 and g , there exists a unique nonholonomic connection ∇ with the properties [Dragovic and Gajic (2003)]

$$\nabla_X g(Y, Z) = X(g(Y, Z)) - g(\nabla_X Y, Z) - g(Y, \nabla_X Z) = 0, \quad T_\nabla = 0.$$

These conditions can be rewritten in the form:

$$\nabla_X Y = \nabla_Y X + p_0[X, Y], \quad Z(g(X, Y)) = g(\nabla_Z X, Y) + g(X, \nabla_Z Y).$$

By cyclic permutation of X, Y, Z and summing we get:

$$g(\nabla_X Y, Z) = \frac{1}{2} \{X(g(Y, Z)) + Y(g(Z, X)) - Z(g(X, Y)) + g(Z, p_0[X, Y]) + g(Y, p_0[Z, X]) - g(X, p_0[Y, Z])\}. \quad (4.42)$$

Let q^i , ($i = 1, \dots, n$) be local coordinates on M , such that the first m coordinate vector-fields ∂_{q^j} are projected by projection p_0 into vector-fields e_a , ($a = 1, \dots, m$), generating the distribution V : $p_0 \partial_{q^j} = p_i^a(q) e_a$. Vector-fields e_a can be expressed in the basis ∂_{q^j} as $e_a = B_a^i \partial_{q^i}$, with $B_a^i p_i^b = \delta_a^b$. Now we give coordinate expressions for the coefficients of the connection Γ_{ab}^c , defined as $\nabla_{e_a} e_b = \Gamma_{ab}^c e_c$. From (4.42) we get

$$\Gamma_{ab}^c = \{^c_{ab}\} + g_{ae} g^{cd} \Omega_{bd}^e + g_{be} g^{cd} \Omega_{ad}^e - \Omega_{ab}^c,$$

where Ω is get from $p_0[e_a, e_b] = -2\Omega_{ab}^c e_c$ as

$$2\Omega_{ab}^c = p_i^c e_a(B_b^i) - p_i^c e_b(B_a^i),$$

and $\{^c_{ab}\} = \frac{1}{2} g^{ce} (e_a(g_{be}) + e_b(g_{ae}) - e_e(g_{ab}))$.

4.9 Application: Geometrical Nonlinear Control

4.9.1 Introduction to Geometrical Nonlinear Control

In this section we give a brief introduction to *geometrical nonlinear control systems*. Majority of techniques developed under this name consider the so-called *affine nonlinear MIMO-systems* of the form (see [Isidori (1989); Nijmeijer and van der Schaft (1990); Lewis (1995); Lewis and Murray (1997); Lewis (1998)])

$$\dot{x}(t) = f_0(x(t)) + u^i(t) f_i(x(t)), \quad (i = 1, \dots, m) \quad (4.43)$$

where $t \mapsto x(t)$ is a curve in a system's *state manifold* M . The vector-field f_0 is called the *drift vector-field*, describing the dynamics of the system in the absence of controls, and the vector-fields f_1, \dots, f_m are the *input vector-fields* or *control vector-fields*, indicating how we are able to actuate the system. The vector-fields f_0, f_1, \dots, f_m are assumed to be real analytic. We do not ask for any sort of linear independence of the control vector-fields f_1, \dots, f_m . We shall suppose that the controls $u : [0, T] \rightarrow U$ are locally integrable with U some subset of \mathbb{R}^m . We allow the length T of the interval on which the control is defined to be arbitrary. It is convenient to denote by $\tau(u)$ the right endpoint of the interval for a given control

u . For a fixed U we denote by \mathcal{U} the collection of all measurable controls taking their values in U . To be concise about this, a *control affine system* is a triple $\Sigma = (M, \mathcal{F} = \{f_0, f_1, \dots, f_m\}, U)$, with all objects as defined above. A *controlled trajectory* for Σ is a pair (c, u) , where $u \in \mathcal{U}$ and where $c : [0, \tau(u)] \rightarrow M$ is defined so that

$$\dot{c}(t) = f_0(c(t)) + u^i(t)f_i(c(t)).$$

One can show that for admissible controls, the curve c will exist at least for sufficiently small times, and that the initial condition $c(0) = x_0$ uniquely defines c on its domain of definition.

For $x \in M$ and $T > 0$ we define several types of *reachable sets* as:

$$\mathcal{R}_\Sigma(x, T) = \{c(T) :$$

$$(c, u) \text{ is a controlled trajectory for } \Sigma \text{ with } \tau(u) = T \text{ and } c(0) = x\},$$

$$\mathcal{R}_\Sigma(x, \leq T) = \cup_{t \in [0, T]} \mathcal{R}_\Sigma(x, t), \quad \mathcal{R}_\Sigma(x) = \cup_{t \geq 0} \mathcal{R}_\Sigma(x, t),$$

that allow us to give several definitions of controllability as follows. Let $\Sigma = (M, \mathcal{F}, U)$ be a control affine system and let $x \in M$. We say that:

- (1) Σ is *accessible* from x if $\text{int}(\mathcal{R}_\Sigma(x)) \neq \emptyset$.
- (2) Σ is *strongly accessible* from x if $\text{int}(\mathcal{R}_\Sigma(x, T)) \neq \emptyset$ for each $T > 0$.
- (3) Σ is *locally controllable* from x if $x \in \text{int}(\mathcal{R}_\Sigma(x))$.
- (4) Σ is *small-time locally controllable* (STLC) from x if there exists $T > 0$ so that $x \in \text{int}(\mathcal{R}_\Sigma(x, \leq T))$ for each $t \in [0, T]$.
- (5) Σ is *globally controllable* from x if $(\mathcal{R}_\Sigma(x)) = M$.

For example, a typical simple system that is accessible but not controllable is given by the following data:

$$M = \mathbb{R}^2, \quad m = 1, \quad U = [-1, 1],$$

$$\dot{x} = u, \quad \dot{y} = x^2.$$

This system is (not obviously) accessible from $(0, 0)$, but is (obviously) not locally controllable from that same point. Note that although $\mathcal{R}_\Sigma((0, 0), \leq T)$ has nonempty interior, the initial point $(0, 0)$ is not in that interior. Thus this is a system that is not controllable in any sense. Note that the system is also strongly accessible.

4.9.2 Feedback Linearization

Recall that the core of control theory is the idea of the *feedback*. In case of nonlinear control, this implies *feedback linearization*.

Exact Feedback Linearization

The idea of feedback linearization is to algebraically transform the nonlinear system dynamics into a fully or partly linear one so that the linear control techniques can be applied. Note that this is not the same as a conventional linearization using Jacobians. In this subsection we will present the modern, geometrical, Lie-derivative based techniques for exact feedback linearization of nonlinear control systems.

The Lie Derivative and Lie Bracket in Control Theory. Recall (see (3.7) above) that given a scalar function $h(x)$ and a vector-field $f(x)$, we define a new scalar function, $\mathcal{L}_f h = \nabla h f$, which is the Lie derivative of h w.r.t. f , i.e., the directional derivative of h along the direction of the vector f . Repeated Lie derivatives can be defined recursively:

$$\mathcal{L}_f^0 h = h, \quad \mathcal{L}_f^i h = \mathcal{L}_f \left(\mathcal{L}_f^{i-1} h \right) = \nabla \left(\mathcal{L}_f^{i-1} h \right) f, \quad (\text{for } i = 1, 2, \dots)$$

Or given another vector-field, g , then $\mathcal{L}_g \mathcal{L}_f h(x)$ is defined as

$$\mathcal{L}_g \mathcal{L}_f h = \nabla (\mathcal{L}_f h) g.$$

For example, if we have a control system

$$\dot{x} = f(x), \quad y = h(x),$$

with the state $x = x(t)$ and the output y , then the derivatives of the output are:

$$\dot{y} = \frac{\partial h}{\partial x} \dot{x} = \mathcal{L}_f h, \quad \text{and} \quad \ddot{y} = \frac{\partial \mathcal{L}_f h}{\partial x} \dot{x} = \mathcal{L}_f^2 h.$$

Also, recall that the curvature of two vector-fields, g_1, g_2 , gives a non-zero Lie bracket, $[g_1, g_2]$ ((3.7.2) see Figure 4.2). Lie bracket motions can generate new directions in which the system can move.

In general, the Lie bracket of two vector-fields, $f(x)$ and $g(x)$, is defined by

$$[f, g] = \text{Ad}_f g = \nabla g f - \nabla f g = \frac{\partial g}{\partial x} f - \frac{\partial f}{\partial x} g,$$

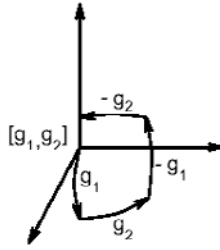


Fig. 4.2 The so-called ‘Lie bracket motion’ is possible by appropriately modulating the control inputs (see text for explanation).

where $\nabla f = \partial f / \partial x$ is the Jacobian matrix. We can define Lie brackets recursively,

$$Ad_f^0 g = g, \quad Ad_f^i g = [f, Ad_f^{i-1} g], \quad (\text{for } i = 1, 2, \dots)$$

Lie brackets have the properties of bilinearity, skew-commutativity and Jacobi identity.

For example, if

$$f = \begin{pmatrix} \cos x_2 \\ x_1 \end{pmatrix}, \quad g = \begin{pmatrix} x_1 \\ 1 \end{pmatrix},$$

then we have

$$[f, g] = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \cos x_2 \\ x_1 \end{pmatrix} - \begin{pmatrix} 0 & -\sin x_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ 1 \end{pmatrix} = \begin{pmatrix} \cos x_2 + \sin x_2 \\ -x_1 \end{pmatrix}.$$

Input/Output Linearization. Given a single-input single-output (SISO) system

$$\dot{x} = f(x) + g(x)u, \quad y = h(x), \quad (4.44)$$

we want to formulate a linear-ODE relation between output y and a new input v . We will investigate (see [Isidori (1989); Sastri and Isidori (1989); Wilson (2000)]):

- How to generate a linear input/output relation.
- What are the internal dynamics and zero-dynamics associated with the input/output linearization?
- How to design stable controllers based on the I/O linearization.

This linearization method will be exact in a finite domain, rather than tangent as in the local linearization methods, which use Taylor series approximation. Nonlinear controller design using the technique is called exact feedback linearization.

Algorithm for Exact Feedback Linearization. We want to find a nonlinear compensator such that the closed-loop system is linear (see Figure 4.3). We will consider only affine SISO systems of the type (4.44), i.e. $\dot{x} = f(x) + g(x)u$, $y = h(x)$, and we will try to construct a *control law* of the form

$$u = p(x) + q(x)v, \quad (4.45)$$

where v is the setpoint, such that the *closed-loop nonlinear system*

$$\dot{x} = f(x) + g(x)p(x) + g(x)q(x)v, \quad y = h(x),$$

is linear from command v to y .

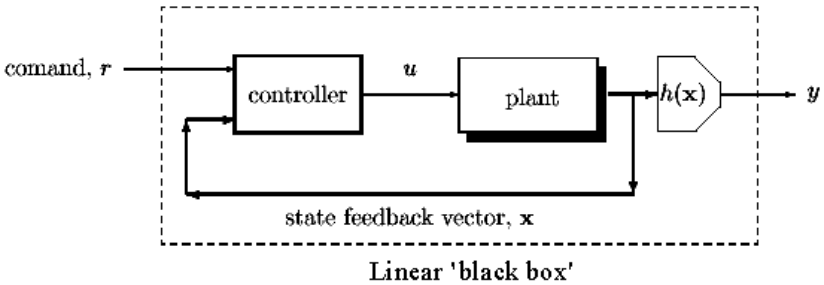


Fig. 4.3 Feedback linearization (see text for explanation).

The main idea behind the feedback linearization construction is to find a nonlinear change of coordinates which transforms the original system into one which is linear and controllable, in particular, a chain of integrators. The difficulty is finding the output function $h(x)$ which makes this construction possible.

We want to design an exact nonlinear feedback controller. Given the nonlinear affine system, $\dot{x} = f(x) + g(x)u$, $y = h(x)$, we want to find the controller functions $p(x)$ and $q(x)$. The unknown functions inside our con-

troller (4.45) are given by:

$$p(x) = \frac{-\left(\mathcal{L}_f^r h(x) + \beta_1 \mathcal{L}_f^{r-1} h(x) + \dots + \beta_{r-1} \mathcal{L}_f h(x) + \beta_r h(x)\right)}{\mathcal{L}_g \mathcal{L}_f^{r-1} h(x)},$$

$$q(x) = \frac{1}{\mathcal{L}_g \mathcal{L}_f^{r-1} h(x)}, \quad (4.46)$$

which are comprised of Lie derivatives, $\mathcal{L}_f h(x)$. Here, the *relative order*, r , is the smallest integer r such that $\mathcal{L}_g \mathcal{L}_f^{r-1} h(x) \neq 0$. For linear systems r is the difference between the number of poles and zeros.

To get the *desired response*, we choose the r parameters in the β polynomial to describe how the output will respond to the setpoint, v (pole-placement).

$$\frac{d^r y}{dt^r} + \beta_1 \frac{d^{r-1} y}{dt^{r-1}} + \dots + \beta_{r-1} \frac{dy}{dt} + \beta_r y = v.$$

Here is the proposed algorithm [Isidori (1989); Sastri and Isidori (1989); Wilson (2000)]:

- (1) Given nonlinear SISO process, $\dot{x} = f(x, u)$, and output equation $y = h(x)$, then:
- (2) Calculate the relative order, r .
- (3) Choose an r th order desired linear response using pole-placement technique (i.e., select β). For this could be used a simple r th order low-pass filter such as a Butterworth filter.
- (4) Construct the exact linearized nonlinear controller (4.46), using Lie derivatives and perhaps a symbolic manipulator (Mathematica or Maple).
- (5) Close the loop and get a linear input-output black-box (see Figure 4.3).
- (6) Verify that the result is actually linear by comparing with the desired response.

Relative Degree

A nonlinear SISO system

$$\dot{x} = f(x) + g(x)u, \quad y = h(x),$$

is said to have *relative degree* r at a point x_o if (see [Isidori (1989); Nijmeijer and van der Schaft (1990)])

- (1) $L_g L_f^k h(x) = 0$ for all x in a neighborhood of x_o and all $k < r - 1$; and
 (2) $L_g L_f^{r-1} h(x_o) \neq 0$.

For example, *controlled Van der Pol oscillator* has the state-space form

$$\dot{x} = f(x) + g(x) u = \begin{bmatrix} x_2 \\ 2\omega\zeta(1 - \mu x_1^2)x_2 - \omega^2 x_1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u.$$

Suppose the output function is chosen as $y = h(x) = x_1$. In this case we have

$$\begin{aligned} L_g h(x) &= \frac{\partial h}{\partial x} g(x) = [1 \ 0] \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0, \quad \text{and} \\ L_f h(x) &= \frac{\partial h}{\partial x} f(x) = [1 \ 0] \begin{bmatrix} x_2 \\ 2\omega\zeta(1 - \mu x_1^2)x_2 - \omega^2 x_1 \end{bmatrix} = x_2. \end{aligned}$$

Moreover

$$L_g L_f h(x) = \frac{\partial(L_f h)}{\partial x} g(x) = [0 \ 1] \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 1,$$

and thus we see that the Vand der Pol oscillator system has relative degree 2 at any point x_o .

However, if the output function is, for instance $y = h(x) = \sin x_2$, then $L_g h(x) = \cos x_2$. The system has relative degree 1 at any point x_o , provided that $(x_o)_2 \neq (2k + 1)\pi/2$. If the point x_o is such that this condition is violated, no relative degree can be defined.

As another example, consider a *linear system* in the state-space form

$$\dot{x} = A x + B u, \quad y = C x.$$

In this case, since $f(x) = A x$, $g(x) = B$, $h(x) = C x$, it can be seen that

$$\begin{aligned} L_f^k h(x) &= C A^k x, \quad \text{and therefore,} \\ L_g L_f^k h(x) &= C A^k B. \end{aligned}$$

Thus, the integer r is characterized by the conditions

$$\begin{aligned} C A^k B &= 0, & \text{for all } k < r - 1 \\ C A^{r-1} B &\neq 0, & \text{otherwise.} \end{aligned}$$

It is well-known that the integer satisfying these conditions is exactly equal to the *difference* between the degree of the denominator polynomial and the

degree of the numerator polynomial of the transfer function

$$H(s) = C(sI - A)^{-1}B$$

of the system.

Approximative Feedback Linearization

Consider a SISO system

$$\dot{x} = f(x) + g(x)u, \quad (4.47)$$

where f and g are smooth vector-fields defined on a compact contractible region M of \mathbb{R}^n containing the origin. (Typically, M is a closed ball in \mathbb{R}^n .) We assume that $f(0) = 0$, i.e., that the origin is an equilibrium for $\dot{x} = f(x)$. The classical problem of feedback linearization can be stated as follows: find in a neighborhood of the origin a smooth change of coordinates $z = \Phi(x)$ (a local diffeomorphism) and a smooth feedback law $u = k(x) + l(x)u_{new}$ such that the closed-loop system in the new coordinates with new control is linear,

$$\dot{z} = Az + Bu_{new},$$

and controllable (see [Banaszuk and Hauser (1996)]). We usually require that $\Phi(0) = 0$. We assume that the system (4.47) has the *linear controllability* property

$$\dim(\text{span}\{g, \text{Ad}_f g, \dots, \text{Ad}_f^{n-1} g\}) = n, \quad \text{for all } x \in M \quad (4.48)$$

(where Ad_f^i are iterated Lie brackets of f and g). We define the *characteristic distribution* for (4.47)

$$\mathcal{D} = \text{span}\{g, \text{Ad}_f g, \dots, \text{Ad}_f^{n-2} g\},$$

which is an $(n-1)$ D smooth distribution by assumption of linear controllability (4.48). We call any nowhere vanishing 1-form ω annihilating \mathcal{D} a characteristic 1-form for (4.47). All the characteristic 1-forms for (4.47) can be represented as multiples of some fixed characteristic 1-form ω_0 by a smooth nowhere vanishing function (zero-form) β . Suppose that there is a non-vanishing β so that $\beta\omega_0$ is exact, i.e., $\beta\omega_0 = d\alpha$ for some smooth function α , where d denotes the exterior derivative. Then ω_0 is called *integrable* and is called an integrating factor for ω_0 . The following result is standard in nonlinear control: Suppose that the system (4.47) has the linear controllability property (4.48) on M . Let \mathcal{D} be the characteristic distribution

and ω_0 be a characteristic 1-form for (4.47). The following statements are equivalent:

- (1) Equation (4.47) is feedback linearizable in a neighborhood of the origin in M ;
- (2) D is involutive in a neighborhood of the origin in M ; and
- (3) ω_0 is integrable in a neighborhood of the origin in M .

As is well known, a generic nonlinear system is not feedback linearizable for $n > 2$. However, in some cases, it may make sense to consider *approximate feedback linearization*.

Namely, if one can find a feedback linearizable system close to (4.47), there is hope that a control designed for the feedback linearizable system and applied to (4.47) will give satisfactory performance if the feedback linearizable system is close enough to (4.47). The first attempt in this direction goes back to [Krener (1984)], where it was proposed to apply to (4.47) a change of variables and feedback that yield a system of the form

$$\dot{z} = Az + B u_{new} + O(z, u_{new}),$$

where the term $O(z, u_{new})$ contains higher-order terms. The aim was to make $O(z, u_{new})$ of as high order as possible. Then we can say that the system (4.47) is approximately feedback linearized in a small neighborhood of the origin. Later [Hunt and Turi (1993)] introduced a new algorithm to achieve the same goal with fewer steps.

Another idea has been investigated in [Hauser *et al.* (1992)]. Roughly speaking, the idea was to neglect nonlinearities in (4.47) responsible for the failure of the involutivity condition in above Theorem. This approach happened to be successful in the ball-and-beam system, when neglect of centrifugal force acting on ball yielded a feedback linearizable system. Application of a control scheme designed for the system with centrifugal force neglected to the original system gave much better results than applying a control scheme based on classical Jacobian linearization. This approach has been further investigated in [Xu and Hauser (1994); Xu and Hauser (1995)] for the purpose of approximate feedback linearization about the manifold of constant operating points. However, a general approach to deciding which nonlinearities should be neglected to get the best approximation has not been set forth.

All of the above-mentioned work dealt with applying a change of coordinates and a preliminary feedback so that the resulting system looks like

linearizable part plus nonlinear terms of highest possible order around an equilibrium point or an equilibrium manifold. However, in many applications one requires a large region of operation for the nonlinearizable system. In such a case, demanding the nonlinear terms to be neglected to be of highest possible order may, in fact, be quite undesirable. One might prefer that the nonlinear terms to be neglected be small in a uniform sense over the region of operation. In this section we propose an approach to approximate feedback linearization that uses a change of coordinates and a preliminary feedback to put a system (4.47) in a perturbed Brunovsky form,

$$\dot{z} = Az + B u_{new} + P(z) + Q(z) u_{new}, \quad (4.49)$$

where $P(z)$ and $Q(z)$ vanish at $z = 0$ and are ‘small’ on M . We get upper bounds on uniform norms of P and Q (depending on some measures of noninvolutivity of \mathcal{D}) on any compact, contractible M .

A different, indirect approach was presented in [Banaszuk and Hauser (1996)]. In this section, the authors present an approach for finding feedback linearizable systems that approximate a given SISO nonlinear system on a given compact region of the state-space. First, they it is shown that if the system is close to being involutive, then it is also close to being linearizable. Rather than working directly with the characteristic distribution of the system, the authors work with characteristic 1-forms, i.e., with the 1-forms annihilating the characteristic distribution. It is shown that homotopy operators can be used to decompose a given characteristic 1-form into an exact and an antiexact part. The exact part is used to define a change of coordinates to a normal form that looks like a linearizable part plus nonlinear perturbation terms. The nonlinear terms in this normal form depend continuously on the antiexact part, and they vanish whenever the antiexact part does. Thus, the antiexact part of a given characteristic 1-form is a measure of nonlinearizability of the system. If the nonlinear terms are small, by neglecting them we get a linearizable system approximating the original system. One can design control for the original system by designing it for the approximating linearizable system and applying it to the original one. We apply this approach for design of locally stabilizing feedback laws for nonlinear systems that are close to being linearizable.

Let us start with approximating characteristic 1-forms by exact forms using *homotopy operators* (compare with equation (3.44) above). Namely, on any contractible region M one can define a linear operator H that sat-

isfies

$$\omega = d(H\omega) + Hd\omega \quad (4.50)$$

for any form ω . The homotopy identity (4.50) allows to decompose any given 1-form into the *exact part* $d(H\omega)$ and an ‘error part’ $\epsilon = Hd\omega$, which we call the *antiexact part of* ω . For given ω_0 annihilating \mathcal{D} and a scaling factor β we define $\alpha_\beta = H\beta\omega_0$ and $\epsilon_\beta = Hd\beta\omega_0$. The 1-form ϵ_β measures how exact $\omega_\beta = \beta\omega_0$ is. If it is zero, then ω_β is exact and the system (4.47) is linearizable, and the zero-form α_β and its first $n - 1$ Lie derivatives along f are the new coordinates. In the case that ω_0 is not exactly integrable, i.e., when no exact integrating factor β exists, we choose β so that $d\beta\omega_0$ is *smallest* in some sense (because this also makes ϵ_β small). We call this β an *approximate integrating factor for* ω_0 . We use the zero-form α_β and its first $n - 1$ Lie derivatives along f as the new coordinates as in the linearizable case. In those new coordinates the system (4.47) is in the form

$$\dot{z} = Az + Bru + Bp + Eu,$$

where r and p are smooth functions, $r \neq 0$ around the origin, and the term E (the obstruction to linearizability) depends linearly on ϵ_β and some of its derivatives. We choose $u = r^{-1}(u_{new} - p)$, where u_{new} is a new control variable. After this change of coordinates and control variable the system is of the form (4.49) with $Q = r^{-1}E$, $P = -r^{-1}pE$. We get estimates on the uniform norm of Q and P (via estimates on r , p , and E) in terms of the error 1-form ϵ_β , for any fixed β , on any compact, contractible manifold M . Most important is that Q and P depend in a continuous way on ϵ_β and some of its derivatives, and they vanish whenever ϵ does (see [Banaszuk and Hauser (1996)]).

4.9.3 Nonlinear Controllability

Linear Controllability

Recall that a system is said to be *controllable* if the set of all states it can reach from initial state $x_0 = x(0)$ at the fixed time $t = T$ contains a ball \mathcal{B} around x_0 . Again, a system is called *small time locally controllable* (STLC) iff the ball \mathcal{B} for $t \leq T$ contains a neighborhood of x_0 .¹³

¹³The above definition of controllability tells us only whether or not something can reach an open neighborhood of its starting point, but does not tell us how to do it. That

In the case of a linear system in the standard state-space form (see subsection (3.13.4.3) above)

$$\dot{x} = Ax + Bu, \quad (4.51)$$

where A is the $n \times n$ *state matrix* and B is the $m \times n$ *input matrix*, all controllability definitions coincide, i.e.,

$$0 \rightarrow x(T), \quad x(0) \rightarrow 0, \quad x(0) \rightarrow x(T),$$

where T is either fixed or free.

Rank condition states: System (4.51) is controllable iff the matrix

$$W_n = (B \ AB \ \dots \ A^{n-1}B) \quad \text{has full rank.}$$

In the case of nonlinear systems the corresponding result is get using the formalism of Lie brackets, as Lie algebra is to nonlinear systems as matrix algebra is to linear systems.

Nonlinear Controllability

Nonlinear MIMO-systems are generally described by differential equations of the form (see [Isidori (1989); Nijmeijer and van der Schaft (1990); Goodwine (1998)]):

$$\dot{x} = f(x) + g_i(x)u^i, \quad (i = 1, \dots, n), \quad (4.52)$$

defined on a smooth n -manifold M , where $x \in M$ represents the state of the control system, $f(x)$ and $g_i(x)$ are vector-fields on M and the u^i are control inputs, which belong to a set of *admissible controls*, $u^i \in U$. The system (4.52) is called *driftless*, or *kinematic*, or *control linear* if $f(x)$ is identically zero; otherwise, it is called a *system with drift*, and the vector-field $f(x)$ is called the *drift term*. The flow $\phi_t^g(x_0)$ represents the solution of the differential equation $\dot{x} = g(x)$ at time t starting from x_0 . Geometrical way to understand the *controllability* of the system (4.52) is to understand the geometry of the vector-fields $f(x)$ and $g_i(x)$.

Example: Car-Parking Using Lie Brackets In this popular example, the driver has two different transformations at his disposal. He/she can turn the steering wheel, or he/she can drive the car forward or back. Here, we specify the state of a car by four coordinates: the (x, y) coordinates is the point of the *trajectory generation*.

of the center of the rear axle, the direction θ of the car, and the angle ϕ between the front wheels and the direction of the car. L is the constant length of the car. Therefore, the configuration manifold of the car is 4D, $M = (x, y, \theta, \phi)$.

Using (4.52), the driftless car kinematics can be defined as:

$$\dot{x} = g_1(x) u_1 + g_2(x) u_2, \quad (4.53)$$

with two vector-fields $g_1, g_2 \in \mathcal{X}^k(M)$.

The infinitesimal transformations will be the vector-fields

$$g_1(x) \equiv \text{DRIVE} = \cos \theta \frac{\partial}{\partial x} + \sin \theta \frac{\partial}{\partial y} + \frac{\tan \phi}{L} \frac{\partial}{\partial \theta} \equiv \begin{pmatrix} \cos \theta \\ \sin \theta \\ \frac{1}{L} \tan \phi \\ 0 \end{pmatrix},$$

$$\text{and } g_2(x) \equiv \text{STEER} = \frac{\partial}{\partial \phi} \equiv \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Now, STEER and DRIVE do not commute; otherwise we could do all your steering at home before driving of on a trip. Therefore, we have a Lie bracket

$$[g_2, g_1] \equiv [\text{STEER}, \text{DRIVE}] = \frac{1}{L \cos^2 \phi} \frac{\partial}{\partial \theta} \equiv \text{ROTATE}.$$

The operation $[g_2, g_1] \equiv \text{ROTATE} \equiv [\text{STEER}, \text{DRIVE}]$ is the infinitesimal version of the sequence of transformations: steer, drive, steer back, and drive back, i.e.,

$$\{\text{STEER}, \text{DRIVE}, \text{STEER}^{-1}, \text{DRIVE}^{-1}\}.$$

Now, ROTATE can get us out of some parking spaces, but not tight ones: we may not have enough room to ROTATE out. The usual tight parking space restricts the DRIVE transformation, but not STEER. A truly tight parking space restricts STEER as well by putting your front wheels against the curb.

Fortunately, there is still another commutator available:

$$\begin{aligned} [g_1, [g_2, g_1]] &\equiv [\text{DRIVE}, [\text{STEER}, \text{DRIVE}]] = [[g_1, g_2], g_1] \equiv \\ [\text{DRIVE}, \text{ROTATE}] &= \frac{1}{L \cos^2 \phi} \left(\sin \theta \frac{\partial}{\partial x} - \cos \theta \frac{\partial}{\partial y} \right) \equiv \text{SLIDE}. \end{aligned}$$

The operation $[[g_1, g_2], g_1] \equiv \text{SLIDE} \equiv [\text{DRIVE}, \text{ROTATE}]$ is a displacement at right angles to the car, and can get us out of any parking place. We just need to remember to steer, drive, steer back, drive some more, steer, drive back, steer back, and drive back:

$$\{\text{STEER}, \text{DRIVE}, \text{STEER}^{-1}, \text{DRIVE}, \text{STEER}, \text{DRIVE}^{-1}, \text{STEER}^{-1}, \text{DRIVE}^{-1}\}.$$

We have to reverse steer in the middle of the parking place. This is not intuitive, and no doubt is part of the problem with parallel parking.

Thus from only two controls u_1 and u_2 we can form the vector-fields $\text{DRIVE} \equiv g_1$, $\text{STEER} \equiv g_2$, $\text{ROTATE} \equiv [g_2, g_1]$, and $\text{SLIDE} \equiv [[g_1, g_2], g_1]$, allowing us to move anywhere in the configuration manifold M . The car kinematics $\dot{x} = g_1 u_1 + g_2 u_2$ is thus expanded as:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \\ \dot{\phi} \end{pmatrix} = \text{DRIVE} \cdot u_1 + \text{STEER} \cdot u_2 \equiv \begin{pmatrix} \cos \theta \\ \sin \theta \\ \frac{1}{L} \tan \phi \\ 0 \end{pmatrix} \cdot u_1 + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \cdot u_2.$$

The *parking Theorem* says: One can get out of any parking lot that is larger than the car.

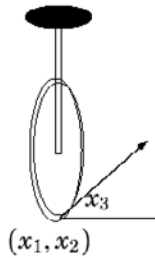


Fig. 4.4 Classical unicycle problem (see text for explanation).

The Unicycle Example. Now, consider the unicycle example (see Figure 4.4). Here we have

$$g_1 = \begin{pmatrix} \cos x_3 \\ \sin x_3 \\ 0 \end{pmatrix}, \quad g_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad [g_1, g_2] = \begin{pmatrix} \sin x_3 \\ -\cos x_3 \\ 0 \end{pmatrix}.$$

The unicycle system is full rank and therefore controllable.

Controllability Condition

Nonlinear controllability is an extension of linear controllability. The nonlinear MIMO system

$$\dot{x} = f(x) + g(x)u \quad \text{is controllable}$$

if the set of vector-fields $\{g, [f, g], \dots, [f^{n-1}, g]\}$ is independent.

For example, for the kinematic car system of the form (4.53), the *non-linear controllability criterion* reads: If the Lie bracket tree:

$$g_1, \quad g_2, \quad [g_1, g_2], \quad [[g_1, g_2], g_1], \quad [[g_1, g_2], g_2], \quad [[[[g_1, g_2], g_1], g_1], \\ [[[[g_1, g_2], g_1], g_2], \quad [[[[g_1, g_2], g_2], g_1], \quad [[[[g_1, g_2], g_2], g_2], \dots$$

– has *full rank* then the system is *controllable* [Isidori (1989); Nijmeijer and van der Schaft (1990); Goodwine (1998)]. In this case the combined input

$$(u_1, u_2) = \begin{cases} (1, 0), & t \in [0, \varepsilon] \\ (0, 1), & t \in [\varepsilon, 2\varepsilon] \\ (-1, 0), & t \in [2\varepsilon, 3\varepsilon] \\ (0, -1), & t \in [3\varepsilon, 4\varepsilon] \end{cases}$$

gives the motion $x(4\varepsilon) = x(0) + \varepsilon^2 [g_1, g_2] + O(\varepsilon^3)$, with the flow given by (see (3.49) below)

$$F_t^{[g_1, g_2]} = \lim_{n \rightarrow \infty} \left(F^{-g_2}_{\sqrt{t/n}} F^{-g_1}_{\sqrt{t/n}} F^{g_2}_{\sqrt{t/n}} F^{g_1}_{\sqrt{t/n}} \right)^n.$$

Distributions

In control theory, the set of all possible directions in which the system can move, or the set of all points the system can reach, is of obvious fundamental importance. Geometrically, this is related to *distributions*.

Recall from subsection 4.7 above that a *distribution* $\Delta \subset \mathcal{X}^k(M)$ on a smooth n D manifold M is a subbundle of its tangent bundle TM , which assigns a subspace of the tangent space $T_x M$ to each point $x \in M$ in a smooth way. The dimension of $\Delta(x)$ over \mathbb{R} at a point $x \in M$ is called the rank of Δ at x .

A distribution Δ is involutive if, for any two vector-fields $X, Y \in \Delta$, their Lie bracket $[X, Y] \in \Delta$.

A function $f \in C^\infty(M)$ is called an *integral* of Δ if $df(x) \in \Delta^0(x)$ for each $x \in M$. An *integral manifold* of Δ is a submanifold N of M such that $T_x N \subset \Delta(x)$ for each $x \in N$. A distribution Δ is *integrable* if, for any $x \in M$, there is a submanifold $N \subset M$, whose dimension is the same

as the rank of Δ at x , containing x such that the tangent bundle, TN , is exactly Δ restricted to N , i.e., $TN = \Delta|_N$. Such a submanifold is called the *maximal integral manifold* through x .

It is natural to consider distributions generated by the vector-fields appearing in the sequence of flows (3.48). In this case, consider the distribution defined by

$$\Delta = \text{span}\{f; g_1 \dots g_m\},$$

where the span is taken over the set of smooth real-valued functions. Denote by $\bar{\Delta}$ the *involutive closure* of the distribution Δ , which is the closure of Δ under bracketing. Then, $\bar{\Delta}$ is the smallest subalgebra of $\mathcal{X}^k(M)$ which contains $\{f; g_1 \dots g_m\}$. We will often need to ‘add’ distributions. Since distributions are, pointwise, vector spaces, define the sum of two distributions,

$$(\Delta_1 + \Delta_2)(x) = \Delta_1(x) + \Delta_2(x).$$

Similarly, define the intersection

$$(\Delta_1 \cap \Delta_2)(x) = \Delta_1(x) \cap \Delta_2(x).$$

More generally, we can arrive at a distribution via a *family of vector-fields*, which is a subset $\mathcal{V} \subset \mathcal{X}^k(M)$. Given a family of vector-fields \mathcal{V} , we may define a distribution on M by

$$\Delta_{\mathcal{V}}(x) = \langle X(x) | X \in \mathcal{V} \rangle_{\mathbb{R}}.$$

Since $\mathcal{X}^k(M)$ is a Lie algebra, we may ask for the smallest Lie subalgebra of $\mathcal{X}^k(M)$ which contains a family of vector-fields \mathcal{V} . It will be denoted as $\overline{\text{Lie}}(\mathcal{V})$, and will be represented by the set of vector-fields on M generated by repeated Lie brackets of elements in \mathcal{V} . Let $\mathcal{V}^{(0)} = \mathcal{V}$ and then iteratively define a sequence of families of vector-fields by

$$\mathcal{V}^{(i+1)} = \mathcal{V}^{(i)} \cup \{[X, Y] | X \in \mathcal{V}^{(0)} = \mathcal{V} \text{ and } Y \in \mathcal{V}^{(i)}\}.$$

Now, every element of $\overline{\text{Lie}}(\mathcal{V})$ is a linear combination of repeated Lie brackets of the form

$$[Z_k, [Z_{k-1}, [\dots, [Z_2, Z_1] \dots]]]$$

where $Z_i \in \mathcal{V}$ for $i = 1, \dots, k$.

Foliations

Recall that related to integrable distributions are *foliations*.

The *Frobenius Theorem* asserts that integrability and involutivity are equivalent, at least locally. Thus, associated with an involutive distribution is a partition Φ of M into disjoint connected immersed submanifolds called *leaves*. This partition Φ is called a *foliation*. More precisely, a foliation \mathcal{F} of a smooth manifold M is a collection of disjoint immersed submanifolds of M whose disjoint union equals M . Each connected submanifold of \mathcal{F} is called a *leaf* of the foliation. Given an integrable distribution Δ , the collection of maximal integral manifolds for Δ defines a foliation on M , denoted by \mathcal{F}_D .

A foliation \mathcal{F} of M defines an equivalence relation on M whereby two points in M are equivalent if they lie in the same leaf of \mathcal{F} . The set of equivalence classes is denoted M/\mathcal{F} and is called the *leaf space* of \mathcal{F} . A foliation \mathcal{F} is said to be simple if M/\mathcal{F} inherits a manifold structure so that the projection from M to M/\mathcal{F} is a surjective submersion.

In control theory, foliation leaves are related to the set of points that a control system can reach starting from a given initial condition. A foliation Φ of M defines an equivalence relation on M whereby two points in M are equivalent if they lie in the same leaf of Φ . The set of equivalence classes is denoted M/Φ and is called the *leaf space* of Φ .

Philip Hall Basis

Given a set of vector-fields $\{g_1 \dots g_m\}$, define the *length* of a *Lie product* as

$$l(g_i) = 1, \quad l([A, B]) = l(A) + l(B), \quad (\text{for } i = 1, \dots, m),$$

where A and B may be Lie products. A *Philip Hall basis* is an ordered set of Lie products $H = \{B_i\}$ satisfying:

- (1) $g_i \in H$, $(i = 1, \dots, m)$;
- (2) If $l(B_i) < l(B_j)$, then $B_i < B_j$; and
- (3) $[B_i, B_j] \in H$ iff
 - (a) $B_i, B_j \in H$ and $B_i < B_j$, and
 - (b) either $B_j = g_k$ for some k or $B_j = [B_l, B_r]$ with $B_l, B_r \in H$ and $B_l \leq B_i$.

Essentially, the ordering aspect of the Philip Hall basis vectors accounts for skew symmetry and Jacobi identity to determine a basis.

4.9.4 Geometrical Control of Mechanical Systems

Much of the existing work on control of mechanical systems has relied on the presence of specific structure. The most common examples of the types of structure assumed are symmetry (conservation laws) and constraints. While it may seem counter-intuitive that constraints may help in control theory, this is sometimes in fact the case. The reason is that the constraints give extra forces (forces of constraint) which can be used to advantage. probably, the most interesting work is done from the Lagrangian (respectively Hamiltonian) perspective where we study systems whose Lagrangians are ‘kinetic energy minus potential energy’ (resp. ‘kinetic energy plus potential energy’). For these *simple mechanical control systems*, the controllability questions are different than those typically asked in nonlinear control theory. In particular, one is often more interested in what happens to configurations rather than states, which are configurations and velocities (resp. momenta) for these systems (see [Lewis (1995); Lewis and Murray (1997)]).

4.9.4.1 Abstract Control System

In general, a nonlinear control system Σ can be represented as a triple (Σ, M, f) , where M is the system’s *state-space* manifold with the tangent bundle TM and the general fibre bundle E , and f is a smooth map, such that the following bundle diagram commutes [Manikonda (1998)]

$$\begin{array}{ccc} E & \xrightarrow{\psi} & TM \\ & \searrow \pi & \swarrow \pi_M \\ & M & \end{array}$$

where $\psi : (x, u) \mapsto (x, f(x, u))$, π_M is the natural projection of TM on M , the projection $\pi : E \rightarrow M$ is a smooth fibre bundle, and the fibers of E represent the *input spaces*. If one chooses fibre-respecting coordinates (x, u) for E , then locally this definition reduces to $\psi : (x, u) \mapsto (x, \psi(x, u))$, i.e.,

$$\dot{x} = \psi(x, u).$$

The specific form of the map ψ , usually used in nonlinear control, is

$\psi : (x, u) \mapsto (x, f(x) + g(x, u))$, with $g(x, 0) = 0$, producing standard nonlinear system equation

$$\dot{x} = f(x) + g(x, u).$$

4.9.4.2 Global Controllability of Linear Control Systems

Consider a linear biodynamical control system:

$$\dot{x}(t) = Ax(t) + Bu(t), \quad (4.54)$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $A \in L(\mathbb{R}^n, \mathbb{R}^n)$, and $B \in L(\mathbb{R}^m, \mathbb{R}^n)$. One should think of $t \mapsto u(t)$ as being a specified input signal, i.e., a function on the certain time interval, $[0, T]$. Now, control theory wants to design the signal to make the *state* $t \mapsto x(t)$ do what we want. What this is may vary, depending on the situation at hand. For example, one may want to steer from an initial state x^i to a final state x_f , perhaps in an optimal way. Or, one may wish to design $u : \mathbb{R}^n \rightarrow \mathbb{R}^m$ so that some state, perhaps $x = 0$, is stable for the dynamical system $\dot{x}(t) = Ax + Bu(x)$, which is called *state feedback* (often one asks that u be linear). One could also design u to be a function of both x and t , etc.

One of the basic control questions is *controllability*, which comes in many guises. Basically we are asking for ‘reachable’ points. In particular,

$$\mathcal{R}(0) = \text{span}_{\mathbb{R}}\{[B|AB|\dots|A^{n-1}B]\},$$

which is the smallest A -invariant subspace containing $\text{Im}(B)$, denotes the set of points reachable from $0 \in \mathbb{R}^n$. For the linear system (4.54), the basic controllability questions have definite answers. We want to do something similar for a class of simple mechanical systems [Lewis (1995); Lewis and Murray (1997)].

4.9.4.3 Local Controllability of Affine Control Systems

The nonlinear control system that we most often consider in humanoid robotics (see next section) has state-space M , a smooth n -manifold, and is *affine* in the controls. Thus it has the form (see [Lewis (1995); Lewis and Murray (1997)])

$$\dot{x} = f(x) + u^a g_a(x), \quad (x \in M), \quad (4.55)$$

where f, g_1, \dots, g_m are vector-fields on M . The *drift* vector-field $f = f(x)$ describes how the system would evolve in the absence of any inputs. Each of the *control* vector-fields g_1, \dots, g_m specifies a direction in which one can supply actuation. To fully specify the control system properly, one should also specify the type of control action to be considered. Here we consider our controls to be taken from the set: $U = \{u : \mathbb{R} \rightarrow \mathbb{R}^m \mid u \text{ is piecewise constant}\}$. This class of controls is sufficient to deal with all analytic control systems. More generally, one may wish to consider measurable functions which take their values in a subset of \mathbb{R}^m .

Given an *affine control system* (4.55), it is possible to define a family of vector-fields on M by: $V_\Sigma = \{f + u^a g_a \mid u \in \mathbb{R}^m\}$.

A solution of the system (4.55) is a pair (γ, u) , where $\gamma : [0, T] \rightarrow M$ is a piecewise smooth curve on M and $u \in U$ such that

$$\dot{\gamma}(t) = f(\gamma(t)) + u^a(t) g_a(\gamma(t)), \quad \text{for each } t \in [0, T].$$

The *reachable set* from x_0 in time T is

$$\begin{aligned} \mathcal{R}(x_0, T) &= \{x \mid \exists \gamma : [0, T] \rightarrow M \quad \text{and} \\ &\quad u : [0, T] \rightarrow \mathbb{R}^m \quad \text{satisfying (4.55)} \\ &\quad \text{with } \gamma(0) = x_0 \quad \text{and } \gamma(T) = x\}. \end{aligned}$$

Note that since the system has drift f , when we reach the point $\gamma(T)$ we will not remain there if this is not an equilibrium point for f . Also, we have, $\mathcal{R}(x_0, \leq T) = \cup_{0 < t \leq T} \mathcal{R}(x_0, T)$.

Let $x_0 \in M$, let V be a neighborhood of x_0 , and let $T > 0$. We say that equation (4.55) represents a *locally accessible system* at x_0 if $\mathcal{R}(x_0, \leq T)$ contains an open subset of M for each V and for each T sufficiently small. Furthermore, we say that the system (4.55) is *small-time local controllability* (STLC, see [Sussmann (1983); Sussmann (1987)]), if it is locally accessible and if x_0 is in the interior of $\mathcal{R}(x_0, \leq T)$ for each V and for each T sufficiently small.

4.9.4.4 Lagrangian Control Systems

Simple Mechanical Control Systems

As a motivation/prototype of a simple mechanical control system, consider a simple *robotic leg* (see Figure 4.5), in which inputs are: (1) an internal torque F^1 moving the leg relative to the body and (2) a force F^2 extending the leg. This system is ‘controllable’ in the sense that, starting

from rest, one can reach any configuration from a given initial configuration. However, as a traditional control system, it is not controllable because of conservation of angular momentum. If one asks for the *states* (i.e., configurations and velocities) reachable from configurations with zero initial velocity, one finds that not all states are reachable. This is a consequence of the fact that angular momentum is conserved, even with inputs. Thus if one starts with zero momentum, the momentum will remain zero (this is what enables one to treat the system as *nonholonomic*). Nevertheless, all configurations are accessible. This suggests that the question of controllability is different depending on whether one is interested in configurations or states. We will be mainly interested in reachable configurations. Considering the system with just one of the two possible input forces is also interesting. In the case where we are just allowed to use F^2 , the possible motions are quite simple; one can only move the ball on the leg back and forth. With just the force F^1 available, things are a bit more complicated. But, for example, one can still say that no matter how you apply the force, the ball will never move ‘inwards’ [Lewis (1995); Lewis and Murray (1997)].

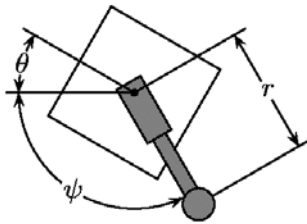


Fig. 4.5 A simple robotic leg (see text for explanation).

In general, simple mechanical control systems are characterized by:

- An n D configuration manifold M ;
- A Riemannian metric g on M ;
- A potential energy function V on M ; and
- m linearly independent 1-forms, F^1, \dots, F^m on M (input forces; e.g., in the case of the simple robotic leg, $F^1 = d\theta - d\psi$ and $F^2 = dr$).

When we say these systems are not amenable to liberalization-based methods, we mean that their liberalizations at zero velocity are not controllable, and that they are not feedback linearizable. This makes simple

mechanical control systems a non-trivial class of nonlinear control systems, especially from the point of view of control design.

As a basic example to start with, consider a *planar rigid body* (see Figure 4.6), with coordinates (x, y, θ) . Inputs are (1) force pointing towards center of mass, $F^1 = \cos \theta dx + \sin \theta dy$, (2) force orthogonal to line to center of mass, $F^2 = -\sin \theta dx + \cos \theta dy - h d\theta$, and (3) torque at center of mass $F^3 = d\theta$. The planar rigid body, although seemingly quite simple, can be actually interesting. Clearly, if one uses all three inputs, the system is *fully actuated*, and so boring for investigating reachable configurations. But if one takes various combinations of one or two inputs, one gets a pretty nice sampling of what can happen for these systems. For example, all possible combinations of two inputs allow one to reach all configurations. Using F^1 or F^3 alone give simple, 1D reachable sets, similar to using F^2 for the robotic leg (as we are always starting with zero initial velocity). However, if one is allowed to only use F^2 , then it is not quite clear what to expect, at least just on the basis of intuition.

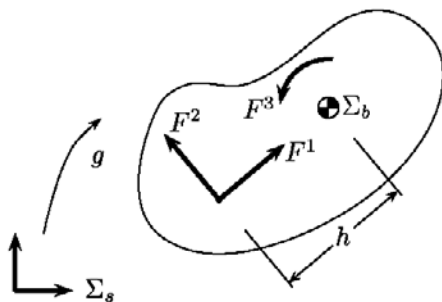


Fig. 4.6 Coordinate systems of a planar rigid body.

It turns out that our simplifying assumptions, i.e., zero initial velocity and restriction of our interest to configurations (i.e., as all problem data is on M , we expect answers to be describable using data on M), makes our task much simpler. In fact, the computations without these assumptions have been attempted, but have yet to yield coherent answers.

Now, we are interested in how do the input 1-forms F^1, \dots, F^m interact with the unforced mechanics of the system as described by the kinetic energy Riemannian metric. That is, what is the analogue of linear system's 'the smallest A -invariant subspace containing $\text{Im}(B)$ ' – for simple mechanical control systems?

Motion and Controllability in Affine Connections

If we start with the local Riemannian metric form $g \mapsto g_{ij}(q) dq^i dq^j$, then we have a kinetic energy Lagrangian $L(q, v) = g_{ij}(q) \dot{q}^i \dot{q}^j$, and consequently the Euler–Lagrangian equations read

$$\frac{d}{dt} \partial_{\dot{q}^i} L - \partial_{q^i} L \equiv g_{ij} \ddot{q}^j + \left(\partial_{q^k} g_{ij} - \frac{1}{2} \partial_{q^i} g_{jk} \right) \dot{q}^j \dot{q}^k = u_a F_i^a, \quad (i = 1, \dots, n).$$

Now multiply this by g^{li} and take the symmetric part of the coefficient of $\dot{q}^j \dot{q}^k$ to get $\ddot{q}^l + \Gamma_{jk}^l \dot{q}^j \dot{q}^k = u^a Y_a^l$, ($l = 1, \dots, n$), where Γ_{jk}^i are the *Christoffel symbols* (3.143) for the *Levi–Civita connection* ∇ (see (3.10.1.1) above). So, the equations of motion can be rewritten

$$\nabla_{\dot{\gamma}(t)} \dot{\gamma}(t) = u^a(t) Y_a(\gamma(t)), \quad (a = 1, \dots, m),$$

where $Y_a = (F^a)^\sharp$, while $^\sharp : T^*M \rightarrow TM$ is the ‘sharp’-isomorphism associated with the Riemannian metric g .

Now, there is nothing to be gained by using a Levi–Civita connection, or by assuming that the vector-fields come from 1-forms. At this point, perhaps the generalization to an arbitrary affine connection seems like a senseless abstraction. However, as we shall see, this abstraction allows us to include another large class of mechanical control systems. So we will study the control system

$$\nabla_{\dot{\gamma}(t)} \dot{\gamma}(t) = u^a(t) Y_a(\gamma(t)) [+Y_0(\gamma(t))], \quad (4.56)$$

with ∇ a general affine connection on M , and Y_1, \dots, Y_m linearly independent vector-fields on M . The ‘optional’ term $Y_0 = Y_0(\gamma(t))$ in (4.56) indicates how potential energy may be added. In this case $Y_0 = -\text{grad } V$ (however, one loses nothing by considering a general vector-field instead of a gradient) [Lewis (1998)].

A *solution* to (4.56) is a pair (γ, u) satisfying (4.56) where $\gamma : [0, T] \rightarrow M$ is a curve and $u : [0, T] \rightarrow \mathbb{R}^m$ is bounded and measurable.

Let U be a neighborhood of $q_0 \in M$ and denote by $\mathcal{R}_M^U(q_0, T)$ those points in M for which there exists a solution (γ, u) with the following properties:

- (1) $\gamma(t) \in U$ for $t \in [0, T]$;
- (2) $\dot{\gamma}(0) = 0_q$; and
- (3) $\gamma(T) \in T_q M$.

Also $\mathcal{R}_M^U(q_0, \leq T) = \cup_{0 \leq t \leq T} \mathcal{R}_M^U(q_0, t)$. Now, regarding the local controllability, we are only interested in points which can be reached without taking ‘large excursions’. Control problems which are local in this way have the advantage that they can be characterized by Lie brackets. So, we want to describe our *reachable set* $\mathcal{R}_M^U(q, \leq T)$ for the simple mechanical control system (4.56). The system (4.56) is *locally configuration accessible* (LCA) at q if there exists $T > 0$ so that $\mathcal{R}_M^U(q, \leq t)$ contains a non-empty open subset of M for each neighborhood U of q and each $t \in]0, T]$. Also, (4.56) is *locally configuration controllable* (LCC) at q if there exists $T > 0$ so that $\mathcal{R}_M^U(q, \leq t)$ contains a neighborhood of q for each neighborhood U of q and each $t \in]0, T]$. Although sound very similar, the notions of local configuration accessibility and local configuration controllability are genuinely different (see Figure 4.7). Indeed, one need only look at the example of the robotic leg with the F^1 input. In this example one may show that the system is LCA, but is not LCC [Lewis (1998)].

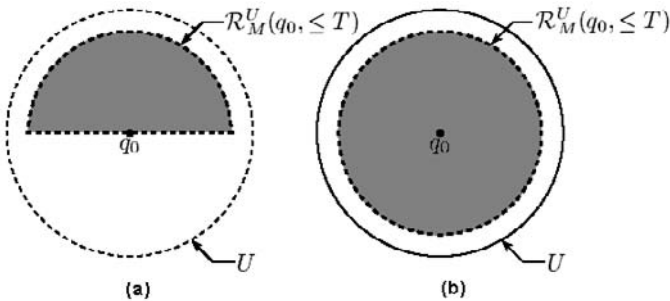


Fig. 4.7 Difference between the notions of local configuration accessibility (a), and local configuration controllability (b).

Local Configuration Accessibility

The accessibility problem is solved by looking at Lie brackets. For this we need to recall the definition of the *vertical lift* [Lewis (1998)]:

$$\text{verlift}(Y(v_q)) = \left. \frac{d}{dt} \right|_{t=0} (v_q + tY(q)),$$

in local coordinates, if $Y = Y^i \partial_{q^i}$, then $\text{verlift}(Y) = Y^i \partial_{v^i}$. Now we can rewrite (4.56) in the first-order form:

$$\dot{v} = Z(v) + u^a \text{verlift}(Y_a(v)),$$

where Z is the geodesic spray for ∇ .

We evaluate all brackets at 0_q (recall that $T_{0_q}TM \simeq T_qM \oplus T_qM$). Here, the first component we think of as being the ‘horizontal’ bit which is tangent to the zero section in TM , and we think of the second component as being the ‘vertical’ bit which is the tangent space to the fibre of $\tau_M : TM \rightarrow M$.

To get an answer to the local configuration accessibility problem, we employ standard nonlinear control techniques involving Lie brackets. Doing so gives us our first look at the *symmetric product*, $\langle X : Y \rangle = \nabla_X Y + \nabla_Y X$. Our sample brackets suggest that perhaps the only things which appear in the bracket computations are symmetric products and Lie brackets of the input vector-fields Y_1, \dots, Y_m .

Here are some sample brackets:

- (i) $[Z, \text{verlift}(Y_a)](0_q) = (-Y_a(q), 0)$;
- (ii) $[\text{verlift}(Y_a), [Z, \text{verlift}(Y_b)]](0_q) = (0, \langle Y_a : Y_b \rangle(q))$;
- (iii) $[[Z, \text{verlift}(Y_a)], [Z, \text{verlift}(Y_b)]](0_q) = ([Y_a, Y_b](q), 0)$.

Now, let C_{ver} be the closure of $\text{span}\{Y_1, \dots, Y_m\}$ under symmetric product. Also, let C_{hor} be the closure of C_{ver} under Lie bracket. So, we assume C_{ver} and C_{hor} to be distributions (i.e., of constant rank) on M . The closure of $\text{span}\{Z, \text{verlift}(Y_1), \dots, \text{verlift}(Y_m)\}$ under Lie bracket, when evaluated at 0_q , is then the distribution

$$q \mapsto C_{hor}(q) \oplus C_{ver}(q) \subset T_qM \oplus T_qM.$$

Proving that the involutive closure of $\text{span}\{Z, \text{verlift}(Y_1), \dots, \text{verlift}(Y_m)\}$ is equal at 0_q to $C_{hor}(q) \oplus C_{ver}(q)$ is a matter of computing brackets, samples of which are given above, and seeing the patterns to suggest an inductive proof. The brackets for these systems are very structured. For example, the brackets of input vector-fields are identically zero. Many other brackets vanish identically, and many more vanish when evaluated at 0_q .

C_{hor} is integrable: let Λ_q be the maximal integral manifold through $q \in M$. Then, $\mathcal{R}_M^U(q, \leq T)$ is contained in Λ_q , and $\mathcal{R}_M^U(q, \leq T)$ contains a non-empty open subset of Λ_q . In particular, if $\text{rank}(C_{hor}) = n$ then (4.56) is LCA [Lewis (1995); Lewis and Murray (1997)]. This Theorem gives a ‘computable’ description of the reachable sets (in the sense that we can calculate Λ_q by solving some over-determined nonlinear PDE’s). But it does not give the kind of insight that we had with the ‘smallest A -invariant subspace containing $\text{Im}(B)$ ’.

Recall that a submanifold N of M is *totally geodesic* if every geodesic

with initial velocity tangent to N remains on N . This can be weakened to distributions: a distribution D on M is *geodesically invariant* if for every geodesic $\gamma : [0, T] \rightarrow M$, $\dot{\gamma}(0) \in D_{\gamma(0)}$ implies $\dot{\gamma}(t) \in D_{\gamma(t)}$ for $t \in]0, T[$.

D is geodesically invariant if it is closed under symmetric product [Lewis (1998)]. This Theorem says that the symmetric product plays for geodesically invariant distributions the same role the Lie bracket plays for integrable distributions. This result was key in providing the geometrical description of the reachable configurations.

An integrable distribution is *geodesically generated distribution* if it is the involutive closure of a geodesically invariant distribution. This basically means that one may reach all points on a leaf with geodesics lying in some subdistribution. The picture one should have in mind with the geometry of the reachable sets is a foliation of M by geodesically generated (immersed) submanifolds onto which the control system restricts if the initial velocity is zero. The idea is that when we start with zero velocity we remain on leaves of the foliation defined by C_{hor} [Lewis and Murray (1997); Lewis (2000a)]. Note that for cases when the affine connection possesses no geodesically invariant distributions, the system (4.56) is automatically LCA. This is true, for example, of S^2 with the affine connection associated with its round metric.

Clearly C_{ver} is the smallest geodesically invariant distribution containing $\text{span}\{Y_1, \dots, Y_m\}$. Also, C_{hor} is geodesically generated by $\text{span}\{Y_1, \dots, Y_m\}$. Thus \mathcal{R}_M^U is contained in, and contains a non-empty open subset of, the distribution geodesically generated by $\text{span}\{Y_1, \dots, Y_m\}$. Note that the pretty decomposition we have for systems with no potential energy does not exist at this point for systems with potential energy.

Local Configuration Controllability

The problem of configuration controllability is harder than the one of configuration accessibility. Following [Lewis and Murray (1999); Lewis (2000a)], we will call a symmetric product in $\{Y_1, \dots, Y_m\}$ *bad* if it contains an even number of each of the input vector-fields. Otherwise we will call it *good*. The *degree* is the total number of vector-fields. For example, $\langle\langle Y_a : Y_b \rangle : Y_a : Y_b \rangle\rangle$ is bad and of degree 4, and $\langle Y_a : \langle Y_b : Y_b \rangle \rangle$ is good and of degree 3. If each bad symmetric product at q is a linear combination of good symmetric products of lower degree, then (4.56) is LCC at q .

Now, the single-input case can be solved completely: The system (4.56) with $m = 1$ is LCC iff $\dim(M) = 1$ [Lewis and Murray (1999)].

Systems With Nonholonomic Constraints

Let us now add to the data a distribution D defining nonholonomic constraints. One of the interesting things about this affine connection approach is that we can easily integrate into our framework systems with nonholonomic constraints. As a simple example, consider a *rolling disk* (see Figure 4.8), with two inputs: (1) a ‘rolling’ torque, $F^1 = d\theta$ and (2) a ‘spinning’ torque, $F^2 = d\phi$. It can be analyzed as a nonholonomic system (see [Lewis (1999); Lewis (2000a)]).

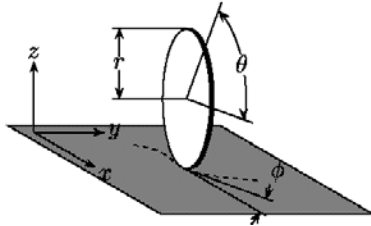


Fig. 4.8 Rolling disk problem (see text for explanation).

The control equations for a simple mechanical control system with constraints are:

$$\nabla_{\dot{\gamma}(t)} \dot{\gamma}(t) = \lambda(t) + u^a(t) Y_a(\gamma(t)) [-\text{grad } V(\gamma(t))], \quad \dot{\gamma}(t) \in D_{\gamma(t)},$$

where $\lambda(t) \in D_{\gamma(t)}^\perp$ are *Lagrangian multipliers*.

Examples

1. Recall that for the simple robotic leg (Figure 4.5) above, Y_1 was internal torque and Y_2 was extension force. Now, in the following three cases:

- (i) both inputs active – this system is LCA and LCC (satisfies sufficient condition);
- (ii) Y_1 only, it is LCA but not LCC; and
- (iii) Y_2 only, it is not LCA.

In these three cases, C_{hor} is generated by the following linearly independent vector-fields:

- (i) both inputs: $\{Y_1, Y_2, [Y_1, Y_2]\}$;
- (ii) Y_1 only: $\{Y_1, \langle Y_1 : Y_1 \rangle, \langle Y_1 : \langle Y_1 : Y_1 \rangle \rangle\}$; and
- (iii) Y_2 only: $\langle Y_2 \rangle$.

Recall that with both inputs the system was not accessible in TM as a consequence of conservation of angular momentum. With the input Y_2 only, the control system behaves very simply when given zero initial velocity. The ball on the end of the leg just gets moved back and forth. This reflects the foliation of M by the maximal integral manifolds of C_{hor} , which are evidently 1D in this case. With the Y_1 input, recall that the ball will always go ‘outwards’ no matter what one does with the input. Thus the system is not LCC. But apparently (since $rank(C_{hor}) = \dim(M)$) one can reach a non-empty open subset of M . The behavior exhibited in this case is typical of what one can expect for single-input systems with no potential energy.

2. For the planar rigid body (Figure 4.6) above, we have the following five cases:

- (i) Y_1 and Y_2 active, this system is LCA and LCC (satisfies sufficient condition);
- (ii) Y_1 and Y_3 , it is LCA and LCC (satisfies sufficient condition);
- (iii) Y_1 only or Y_3 only, not LCA;
- (iv) Y_2 only, LCA but not LCC; and
- (v) Y_2 and Y_3 : LCA and LCC (fails sufficient condition).

Now, with the inputs Y_1 or Y_3 alone, the motion of the system is simple. In the first case the body moves along the line connecting the point of application of the force and the center of mass, and in the other case the body simply rotates. The equations in (x, y, θ) coordinates are

$$\ddot{x} = \frac{\cos \theta}{m} u^1 - \frac{\sin \theta}{m} u^2, \quad \ddot{y} = \frac{\sin \theta}{m} u^1 + \frac{\cos \theta}{m} u^2, \quad \ddot{\theta} = \frac{1}{J} (u^3 - hu^2),$$

which illustrates that the θ -equation decouples when only Y_3 is applied. We make a change of coordinates for the case where we have only Y_1 : $(\xi, \eta, \psi) = (x \cos \theta + y \sin \theta, -x \sin \theta + y \cos \theta, \theta)$. In these coordinates we have

$$\ddot{\xi} - 2\dot{\eta}\dot{\psi} - \xi\dot{\psi}^2 = \frac{1}{m}u^1, \quad \ddot{\eta} + 2\dot{\xi}\dot{\psi} - \eta\dot{\psi}^2 = 0, \quad \dot{\psi} = 0,$$

which illustrates the decoupling of the ξ -equation in this case.

C_{hor} has the following generators:

- (i) Y_1 and Y_2 : $\{Y_1, Y_2, [Y_1, Y_2]\}$;
- (ii) Y_1 and Y_3 : $\{Y_1, Y_3, [Y_1, Y_3]\}$;
- (iii) Y_1 only or Y_3 only: $\{Y_1\}$ or $\{Y_3\}$;
- (iv) Y_2 only: $\{Y_2, \langle Y_2 : Y_2 \rangle, \langle Y_2 : \langle Y_2 : Y_2 \rangle \rangle\}$;
- (v) Y_2 and Y_3 $\{Y_2, Y_3, [Y_2, Y_3]\}$.

3. Recall that for the rolling disk (Figure 4.8) above, Y_1 was ‘rolling’ input and Y_2 was ‘spinning’ input. Now, in the following three cases:

(i) Y_1 and Y_2 active, this system is LCA and LCC (satisfies sufficient condition);

(ii) Y_1 only: not LCA; and

(iii) Y_2 only: not LCA.

In these three cases, C_{hor} has generators:

(i) Y_1 and Y_2 : $\{Y_1, Y_2, [Y_1, Y_2], [Y_2, [Y_1, Y_2]]\}$;

(ii) Y_1 only: $\{Y_1\}$; and

(iii) Y_2 only: $\{Y_2\}$.

The rolling disk passes the good/bad symmetric product test. Another way to show that it is LCC is to show that the inputs allow one to follow any curve which is admitted by the constraints. Local configuration controllability then follows as the constraint distribution for the rolling disk has an involutive closure of maximal rank [Lewis (1999)].

Categorical Structure of Control Affine Systems

Control affine systems make a category \mathcal{CAS} (see [Elkin (1999)]). The category \mathcal{CAS} has the following data:

- An *object* in \mathcal{CAS} is a pair $\Sigma = (M, \mathfrak{F} = \{f_0, f_1, \dots, f_m\})$ where \mathfrak{F} is a family of vector-fields

$$\dot{x}(t) = f_0(x(t)) + u^a(t)f_a(x(t))$$

on the manifold M .

- A *morphism* sending $\Sigma = (M, \mathfrak{F} = \{f_0, f_1, \dots, f_m\})$ to $\Sigma' = (M', \mathfrak{F}' = \{f'_0, f'_1, \dots, f'_{m'}\})$ is a triple $(\psi, \lambda_0, \Lambda)$ where $\psi : M \rightarrow M'$, $\lambda_0 : M \rightarrow \mathbb{R}^{m'}$, and $\Lambda : M \rightarrow L(\mathbb{R}^m, \mathbb{R}^{m'})$ are smooth maps satisfying:

- (1) $T_x\psi(f_a(x)) = \Lambda_a^\alpha(x)f'_\alpha(\psi(x))$, $a \in \{1, \dots, m\}$, and
- (2) $T_x\psi(f_0(x)) = f'_0(\psi(x)) + \lambda_0^\alpha f'_\alpha(\psi(x))$.

This corresponds to a change of state-input by

$$(x, u) \longmapsto (\psi(x), \lambda_0(x) + \Lambda(x)u).$$

Elkin [Elkin (1999)] discusses equivalence, inclusion, and factorization in the category \mathcal{CAS} . Using categorical language, he considers local equivalence for various classes of nonlinear control systems, including single-input systems, systems with involutive input distributions, and systems with three states and two inputs.

4.9.4.5 Lie-Adaptive Control

In this subsection we develop the concept of *machine learning* in the framework of Lie derivative control formalism (see (4.9.2) above). Consider an n D, SISO system in the standard affine form (4.44), rewritten here for convenience:

$$\dot{x}(t) = f(x) + g(x)u(t), \quad y(t) = h(x), \quad (4.57)$$

As already stated, the feedback control law for the system (4.57) can be defined using Lie derivatives $\mathcal{L}_f h$ and $\mathcal{L}_g h$ of the system's output h along the vector-fields f and g .

If the SISO system (4.57) is a relatively simple (quasilinear) system with *relative degree* $r = 1$ it can be rewritten in a quasilinear form

$$\dot{x}(t) = \gamma_i(t) f_i(x) + d_j(t) g_j(x) u(t), \quad (4.58)$$

where γ_i ($i = 1, \dots, n$) and d_j ($j = 1, \dots, m$) are system's parameters, while f_i and g_j are smooth vector-fields.

In this case the feedback control law for *tracking the reference signal* $y_R = y_R(t)$ is defined as (see [Isidori (1989); Nijmeijer and van der Schaft (1990)])

$$u = \frac{-\mathcal{L}_f h + \dot{y}_R + \alpha(y_R - y)}{\mathcal{L}_g h}, \quad (4.59)$$

where α denotes the feedback gain.

Obviously, the problem of reference signal tracking is relatively simple and straightforward if we know all the system's parameters $\gamma_i(t)$ and $d_j(t)$ of (4.58). The question is can we apply a similar control law if the system parameters are unknown?

Now we have much harder problem of *adaptive signal tracking*. However, it appears that the feedback control law can be actually cast in a similar form (see [Sastri and Isidori (1989); Gómez (1994)]):

$$\hat{u} = \frac{-\widehat{\mathcal{L}_f h} + \dot{y}_R + \alpha(y_R - y)}{\widehat{\mathcal{L}_g h}}, \quad (4.60)$$

where Lie derivatives $\mathcal{L}_f h$ and $\mathcal{L}_g h$ of (4.59) have been replaced by their estimates $\widehat{\mathcal{L}_f h}$ and $\widehat{\mathcal{L}_g h}$, defined respectively as

$$\widehat{\mathcal{L}_f h} = \hat{\gamma}_i(t) \mathcal{L}_{f_i} h, \quad \widehat{\mathcal{L}_g h} = \hat{d}_j(t) \mathcal{L}_{g_j} h,$$

in which $\hat{\gamma}_i(t)$ and $\hat{d}_j(t)$ are the estimates for $\gamma_i(t)$ and $d_j(t)$.

Therefore, we have the straightforward control law even in the uncertain case, provided that we are able to estimate the unknown system parameters. Probably the best known *parameter update law* is based on the so-called *Lyapunov criterion* (see [Sastri and Isidori (1989)]) and given by

$$\dot{\psi} = -\gamma \epsilon W, \quad (4.61)$$

where $\psi = \{\gamma_i - \hat{\gamma}_i, d_j - \hat{d}_j\}$ is the parameter estimation error, $\epsilon = y - y_R$ is the output error, and γ is a positive constant, while the matrix W is defined as:

$$W = [W_1^T \ W_2^T]^T, \quad \text{with}$$

$$W_1 = \begin{bmatrix} \mathcal{L}_{f_1} h \\ \vdots \\ \mathcal{L}_{f_n} h \end{bmatrix}, \quad W_2 = \begin{bmatrix} \mathcal{L}_{g_1} h \\ \vdots \\ \mathcal{L}_{g_m} h \end{bmatrix} \cdot \frac{-\widehat{\mathcal{L}_f h} + \dot{y}_R + \alpha(y_R - y)}{\widehat{\mathcal{L}_g h}}.$$

The proposed adaptive control formalism (4.60–4.61) can be efficiently applied wherever we have a problem of tracking a given signal with an output of a SISO-system (4.57–4.58) with unknown parameters.

4.9.5 *Hamiltonian Optimal Control and Maximum Principle*

4.9.5.1 *Hamiltonian Control Systems*

Hamiltonian control system on a symplectic manifold (P, ω) is defined as an affine control system whose drift and control vector-fields are Hamiltonian. It can be written as

$$\dot{p} = X_H(p) + u^a X_a(p),$$

where the vector-fields X_a are assumed to be Hamiltonian with Hamiltonian H_a for $a = 1, \dots, m$. Examples of systems which are (at least locally) Hamiltonian control systems are those which evolve on the symplectic manifold T^*M and where the control Hamiltonians are simply coordinate functions on M .

Alternatively, Hamiltonian control systems can be defined on Poisson manifolds. However, for the purposes of this subsection, it will be more natural to work within the Poisson context. Recall that given a smooth Hamiltonian function $h : M \rightarrow \mathbb{R}$, on the Poisson manifold M , the Poisson bracket $\{, \} : C^\infty(M) \times C^\infty(M) \rightarrow C^\infty(M)$ (such that $\{f, g\} = -\{g, f\}$,

$\{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{f, g\}\} = 0$, and $\{fg, h\} = \{f, h\}g + f\{g, h\}$ allows us to get a Hamiltonian vector-field X_h with Hamiltonian h through the equality

$$\mathcal{L}_{X_h} f = \{f, h\}, \quad \text{for all } f \in C^\infty(M),$$

where $\mathcal{L}_{X_h} f$ is the Lie derivative of f along X_h . Note that the vector-field X_h is well defined since the Poisson bracket verifies the Leibniz rule and therefore defines a derivation on $C^\infty(M)$ (see [Marsden and Ratiu (1999)]). Furthermore $C^\infty(M)$ equipped with a Poisson bracket is a Lie algebra, called a Poisson algebra. Also, we say that the Poisson structure on M is nondegenerate if the $\{, \}$ -associated map $B^\# : T^*M \rightarrow TM$ defined by

$$dg(B^\#(x)(df)) = B(x)(df, dg),$$

(where df denotes the exterior derivative of f) is an isomorphism for every $x \in M$.

An *affine Hamiltonian control system* $\Sigma = (U, M, h)$ consists of a smooth manifold U (the *input space*), a Poisson manifold M with nondegenerate Poisson bracket (the *state-space*), and a smooth function $H : M \times U \rightarrow \mathbb{R}$ (the *controlled Hamiltonian*). Furthermore, H is locally of the form $H = h_0 + h_i u^i$ ($i = 1, \dots, n$), with h_i locally defined smooth real valued maps and u^i local coordinates for U [Tabuada and Pappas (2001)].

Using the controlled Hamiltonian and the Poisson structure on M we can recover the familiar system map $F : M \times U \rightarrow TM$, locally given by

$$F = X_{h_0} + X_{h_i} u^i,$$

and defines an affine distribution on M given by

$$\mathcal{D}_M(x) = X_{h_0}(x) + \text{span}\{X_{h_1}(x), X_{h_2}(x), \dots, X_{h_n}(x)\}.$$

This distribution captures all the possible directions of motion available at a certain point x , and therefore describes a control system, up to a parametrization by control inputs. This affine distribution will be our main object of interest here, and we will assume that the rank of \mathcal{D}_M does not change with x . Furthermore, we denote an affine distribution \mathcal{D}_M by $X + \Delta$, where X is a vector-field and Δ a distribution. When this affine distribution is defined by a Hamiltonian control system we have $X = X_{h_0}$ and $\Delta = \text{span}\{X_{h_1}(x), X_{h_2}(x), \dots, X_{h_n}(x)\}$. A similar reasoning is possible at the level of Hamiltonians. Locally, we can define the following affine space of

smooth maps

$$\mathcal{H}_M = h_0 + \text{span}_{\mathbb{R}}\{h_1, h_1, \dots, h_n\},$$

which defines \mathcal{D}_M by the equality

$$\mathcal{D}_M = B^\#(d\mathcal{H}_M),$$

where we used the notation $d\mathcal{H}_M$ to denote the set $\cup_{h \in \mathcal{H}_M} dh$. We also use the notation $\mathcal{H}_M = h_0 + H_\Delta$ for an affine space of smooth maps where h_0 is a smooth map and H_Δ a linear space of smooth maps.

Having defined Hamiltonian control systems we turn to their trajectories or solutions: A smooth curve $\gamma : I \rightarrow M$, $I \subseteq \mathbb{R}_0^+$ is called a *trajectory of control system* $\Sigma = (U, M, H)$, iff there exists a curve $\gamma^U : I \rightarrow U$ satisfying [Tabuada and Pappas (2001)]

$$\dot{\gamma}(t) = F(\gamma(t), \gamma^U(t)), \quad \text{for every } t \in I.$$

Now, given a Hamiltonian control system and a desired property, an abstracted Hamiltonian system is a reduced system that preserves the property of interest while ignoring modelling detail (see [Tabuada and Pappas (2001)]). Property preserving abstractions of control systems are important for reducing the complexity of their analysis or design. From an analysis perspective, given a large scale control system and a property to be verified, one extracts a smaller abstracted system with equivalent properties. Checking the property on the abstraction is then equivalent to checking the property on the original system. From a design perspective, rather than designing a controller for the original large scale system, one designs a controller for the smaller abstracted system, and then refines the design to the original system while incorporating modelling detail.

This approach critically depends on whether we are able to construct hierarchies of abstractions as well as characterize conditions under which various properties of interest propagate from the original to the abstracted system and vice versa. In [Pappas *et al.* (2000)], hierarchical abstractions of linear control systems were extracted using computationally efficient constructions, and conditions under which controllability of the abstracted system implied controllability of the original system were obtained. This led to extremely efficient hierarchical controllability algorithms. In the same spirit, abstractions of nonlinear control affine systems were considered in [Pappas and Simic (2002)], and the canonical construction for linear systems was generalized to nonlinear control affine systems.

In [Tabuada and Pappas (2001)], abstractions of Hamiltonian control systems are considered, which are control systems completely specified by controlled Hamiltonians. This additional structure allows to simplify the abstraction process by working with functions instead of vector-fields or distributions as is the case for general nonlinear systems [Pappas and Simic (2002)]. This is possible since the controlled Hamiltonian contains all the relevant information that must be captured by the abstracted system. On the other hand, to be able to relate the dynamics induced by the controlled Hamiltonians, we need to restrict the class of abstracting maps to those that preserve the Hamiltonian structure. More precisely, given a Hamiltonian control system on a Poisson manifold M , and a (quotient) Poisson map $\phi : M \rightarrow N$, one presents a canonical construction that extracts an abstracted Hamiltonian control system on N . One then characterizes abstracting maps for which the original and abstracted system are equivalent from a local accessibility point of view [Tabuada and Pappas (2001)].

4.9.5.2 Pontryagin's Maximum Principle

Recall that the *Pontryagin Maximum Principle* (PMP, see [Pontryagin *et al.* (1986); Iyanaga and Kawada (1980)]) applies to a general optimization problem called a *Bolza problem*. To apply PMP to optimal control, we need to define Hamiltonian function:

$$H(\psi, x, u) = (\psi, f(x, u)) = \psi_i f^i(x, u), \quad (i = 1, \dots, n). \quad (4.62)$$

Then in order for a control $u(t)$ and a trajectory $x(t)$ to be *optimal*, it is necessary that there exist a nonzero absolutely continuous vector function $\psi(t) = (\psi_0(t), \psi_1(t), \dots, \psi_n(t))$ corresponding to the functions $u(t)$ and $x(t)$ such that:

- (1) The function $H(\psi(t), x(t), u(t))$ attains its maximum at the point $u = u(t)$ almost everywhere in the interval $t_0 \leq t \leq T$,

$$H(\psi(t), x(t), u(t)) = \max_{u \in U} H(\psi(t), x(t), u(t)).$$

- (2) At the *terminal time* T , the following relations are satisfied:
 $\psi_0(T) \leq 0$ and $H(\psi(T), x(T), u(T)) = 0$.

PMP states the following algorithm: To maximize the set of *steering functions* $\gamma_i x^i(t)$ (with n constants γ_i) for controlling the changes in the

state variables

$$\dot{x}^i(t) = f^i(x^i, u_k), \quad (i = 0, 1, \dots, n, \quad k = 1, \dots, m),$$

we maximize at each instant the Hamiltonian function (4.62), where

$$\dot{\psi}_i = -\psi_j \frac{\partial f^j}{\partial x^i} \quad \text{and} \quad \psi_i(T) = \gamma_i.$$

4.9.5.3 Affine Control Systems

Now, let us look at PMP as applied to the *affine control system* (see [Lewis (2000b)])

$$\dot{\gamma}(t) = f_0(\gamma(t)) + u^a(t) f_a(\gamma(t)),$$

with $\gamma(t) \in M$, u taking values in $U \subset \mathbb{R}^m$, and objective function $L(x, u)$.

We need to have the control Hamiltonian on $U \times T^*M$:

$$H(\alpha_x, u) = \underbrace{\alpha_x(f_0(x))}_{H_1} + \underbrace{\alpha_x(u^a f_a(x))}_{H_2} - \underbrace{L(x, u)}_{H_3}.$$

One of several consequences of the PMP is that if (u, γ) is a *minimizer* then there exists a 1-form field λ along γ with the property that $t \mapsto \lambda(t)$ is an integral curve for the time-dependent Hamiltonian $(\alpha_x, u) \mapsto H(\alpha_x, u)$. The Hamiltonian $H(\alpha_x, u)$ is a sum of three terms, and so too will be the Hamiltonian vector-field.

Let us look at the first term, that with (old) Hamiltonian $H_1 = \alpha_x(f_0(x))$. In local coordinates X_{H_1} is written as

$$\dot{x}^i = f_0^i(x), \quad \dot{p}_i = -\frac{\partial f_0^j(x)}{\partial x^i} p_j. \quad (4.63)$$

X_{H_1} is the *cotangent lift* of f_0 and, following [Lewis (2000b)], we denote it f_0^{T*} . So we want to understand f_0^{T*} on TM with $f_0 = Z$.

Let f_0 be a vector-field on a general manifold N with f_0^T its *tangent lift* defined by

$$f_0^T(v_x) = \left. \frac{d}{dt} \right|_{t=0} T_x F_t(v_x),$$

where F_t denotes the flow of f_0 . Therefore, f_0^T is the ‘linearization’ of f_0 and in local coordinates it is given by (compare with (4.63))

$$\dot{x}^i = f_0^i(x), \quad \dot{v}^i = -\frac{\partial f_0^j(x)}{\partial x^j} v^j.$$

The flow of f_0^T measures how the integral curves of f_0 change as we change the initial condition in the direction of v_x .

Now, perhaps we can understand Z^T on TM with $f_0 = Z$ in the discussion of tangent lift. Let $\gamma(t)$ be a geodesic. By varying the initial condition for the geodesic we generate an ‘infinitesimal variation’ which satisfies the *extended Jacobi equation*,

$$\nabla_{\dot{\gamma}(t)}^2 \xi(t) + R(\xi(t), \dot{\gamma}(t)) \dot{\gamma}(t) + \nabla_{\dot{\gamma}(t)} (T(\xi(t), \dot{\gamma}(t))) = 0. \quad (4.64)$$

To make the ‘connection’ between Z^T and the Jacobi equation, we perform constructions on the tangent bundle using the spray Z . ∇ comes from a linear connection on M which induces an *Ehresmann connection* on $\tau_M : TM \rightarrow M$. Thus we may write $T_{v_q} TM \simeq T_q M \oplus T_q M$. Now, if $I_M : TTM \rightarrow TTM$ is the canonical involution then $I_M^* Z^T$ is a spray. We use $I_M^* Z^T$ to induce an Ehresmann connection on $\tau_{TTM} : TTM \rightarrow TM$. Thus,

$$T_{X_{v_q}} TTM \simeq T_{v_q} TM \oplus T_{v_q} TM \simeq \underbrace{T_q M \oplus T_q M}_{\text{geodesic equations}} \oplus \underbrace{T_q M \oplus T_q M}_{\text{variation equations}}.$$

One represents Z^T in this splitting and determines that the Jacobi equation sits ‘inside’ one of the four components. Now one applies similar constructions to T^*TM and Z^{T*} to derive a 1-form version of the Jacobi equation (4.64), the so-called *adjoint Jacobi equation* [Lewis (2000b)]:

$$\nabla_{\dot{\gamma}(t)}^2 \lambda(t) + R^*(\lambda(t), \dot{\gamma}(t)) \dot{\gamma}(t) - T^* (\nabla_{\dot{\gamma}(t)} \lambda(t), \dot{\gamma}(t)) = 0, \quad (4.65)$$

where we have used $\langle R^*(\alpha, u)v; \omega \rangle = \langle \alpha; R(\omega, u)v \rangle$, and $\langle T^*(\alpha, u); \omega \rangle = \langle \alpha; T(\omega, u) \rangle$.

The adjoint Jacobi equation forms the backbone of a general statement of the PMP for affine connection control systems. When objective function is the Lagrangian $L(u, v_q) = \frac{1}{2}g(v_q, v_q)$, when ∇ is the *Levi-Civita connection* for the Riemannian metric g , and when the system is fully actuated, then we recover the equation of [Noakes *et al.* (1989)]

$$\nabla_{\dot{\gamma}(t)}^3 \dot{\gamma}(t) + R(\nabla_{\dot{\gamma}(t)} \dot{\gamma}(t), \dot{\gamma}(t)) = 0.$$

Therefore, the adjoint Jacobi equation (4.65) captures the interesting part of the Hamiltonian vector-field Z^{T*} , which comes from the PMP, in terms of affine geometry, i.e., from Z^{T*} follows

$$\nabla_{\dot{\gamma}(t)} \dot{\gamma}(t) = 0, \quad \nabla_{\dot{\gamma}(t)}^2 \lambda(t) + R^*(\lambda(t), \dot{\gamma}(t)) \dot{\gamma}(t) - T^* (\nabla_{\dot{\gamma}(t)} \lambda(t), \dot{\gamma}(t)) = 0.$$

The geometry of Z on TM gives a way of globally pulling out the adjoint Jacobi equation from the PMP in an intrinsic manner, which is not generally possible in the PMP [Lewis (2000b)].

4.9.6 Brain-Like Control Functor in Biodynamics

In this final section we propose our most recent model [Ivancevic and Beagley (2005)] of the complete biodynamical *brain-like control functor*. This is a neurodynamical reflection on our *covariant force law*, $F_i = mg_{ij}a^j$, and its associated *covariant force functor* $\mathcal{F}_* : TT^*M \rightarrow TTM$ (see section 3.13.4.1 above).

Recall that traditional hierarchical robot control (see, e.g., [Vukobratovic and Stokic (1982); Vukobratovic *et al.* (1990)]) consists of three levels: the *executive* control-level (at the bottom) performs tracking of nominal trajectories in internal-joint coordinates, the *strategic* control-level (at the top) performs ‘planning’ of trajectories of an end-effector in external-Cartesian coordinates, and the *tactical* control-level (in the middle) connects other two levels by means of inverse kinematics.

The modern version of the hierarchical robot control includes decision-making done by the neural (or, neuro-fuzzy) classifier to adapt the (manipulator) control to dynamically changing environment.

On the other hand, the so-called ‘intelligent’ approach to robot control typically represents a form of function approximation, which is itself based on some combination of neuro-fuzzy-genetic computations. Many special issues and workshops focusing on physiological models for robot control reflect the increased attention for the development of *cerebellar models* [van der Smagt (1999); Schaal and Atkeson (1998); Schaal (1999); Schaal (1998); Arbib (1998)] for learning robot control with functional decomposition, where the main result could be formulated as: *the cerebellum is more than just the function approximator*.

In this section we try to fit between these three approaches for humanoid control, emphasizing the role of muscle-like robot actuators. We propose a new, physiologically based, tensor-invariant, hierarchical force control (FC, for short) for the physiologically realistic biodynamics. We consider the muscular torque one-forms F_i as the most important component of human-like motion; therefore we propose the sophisticated hierarchical system for the subtle F_i —control: corresponding to the spinal, the cerebellar and cortical levels of human motor control. F_i are first set-up as testing input-signals to biodynamics, and then covariantly updated as

feedback 1-forms u_i on each FC-level. On the spinal FC-level the nominal joint-trajectory tracking is proposed in the form of affine Hamiltonian control; here the driving torques are given corrections by spinal-reflex controls. On the cerebellar FC-level, the relation is established between canonical joint coordinates q^i, p_i and gradient *neural-image coordinates* x^i, y_i , representing bidirectional, self-organized, associative memory machine; here the driving torques are given the cerebellar corrections. On the cortical FC-level the topological ‘hyper-joystick’ is proposed as the central FC command-space, selector, with the fuzzy-logic feedback-control map defined on it, giving the cortical corrections to the driving torques.

The model of the spinal FC-level formulated here resembles *autogenetic motor servo*, acting on the spinal-reflex level of the human locomotor control. The model of the cerebellar FC-level formulated here mimics the self-organizing, associative function of the excitatory granule cells and the inhibitory Purkinje cells of the cerebellum [Houk *et al.* (1996)]. The model of the cortical FC-level presented in this section mimics the synergistic *regulation of locomotor conditioned reflexes* by the cerebellum [Houk *et al.* (1996)].

We believe that (already mentioned) extremely high order of the driving force redundancy in biodynamics justifies the formulation of the three-level force control system. Also, both brain-like control systems can be easily extended to give $SE(3)$ -based force control for moving inverse kinematics (IK) chains of legs and arms.

4.9.6.1 Functor Control Machine

In this subsection we define the functor control-machine (compare with section (3.13.4.2) above), for the learning control with functional decomposition, by a two-step generalization of the Kalman’s theory of linear MIMO-feedback systems. The first generalization puts the Kalman’s theory into the pair of mutually dual linear categories **Vect** and **Vect*** of vector spaces and linear operators, with a ‘loop-functor’ representing the closed-loop control, thus formulating the unique, categorical formalism valid both for the discrete and continual MIMO-systems.

We start with the unique, *feedforward continual-sequential state equation*

$$\dot{x}(t+1) = Ax(t) + Bu(t), \quad y(t) = Cx(t), \quad (4.66)$$

where the nD vector spaces of *state* $X \ni x$, *input* $U \ni u$, and *output*

$Y \ni y$ have the corresponding linear operators, respectively $A : X \rightarrow X$, $B : U \rightarrow X$, and $C : X \rightarrow Y$. The modular system theory comprises the *system dynamics*, given by a pair (X, A) , together with a *reachability map* $e : U \rightarrow X$ of the pair (B, A) , and an *observability map* $m : X \rightarrow Y$ of the pair (A, C) . If the reachability map e is surjection the system dynamics (X, A) is called *reachable*; if the observability map m is injection the system dynamics (X, A) is called *observable*. If the system dynamics (X, A) is both reachable and observable, a *composition* $r = m \circ e : U \rightarrow Y$ defines the *total system's response*, which is given by solution of equation (4.66). If the unique solution to the continual-sequential state equation exists, it gives the answer to the (minimal) *realization problem*: find the system S that realizes the given response $r = m \circ e : U \rightarrow Y$ (in the smallest number of discrete states and in the shortest time).

The inverse map $r^{-1} = e^{-1} \circ m^{-1} : Y \rightarrow U$ of the total system's response $r : U \rightarrow Y$ defines the linear *feedback operator* $K : Y \rightarrow U$, given by standard feedback equation

$$u(t) = Ky(t). \quad (4.67)$$

In categorical language, the feedforward system dynamics in the category **Vect** is a pair (X, A) , where $X \in \mathbf{Ob}(\mathbf{Vect})$ is an object in **Vect** and $A : X \rightarrow X \in \mathbf{Mor}(\mathbf{Vect})$ is a **Vect**-morphism. A *feedforward decomposable system* in **Vect** is such a sextuple $S \equiv (X, A, U, B, Y, C)$ that (X, A) is the system dynamics in **Vect**, a **Vect**-morphism $B : U \rightarrow X$ is an *input map*, and a **Vect**-morphism $C : X \rightarrow Y$ is an *output map*. Any object in **Vect** is characterized by mutually dual notions of its *degree* (a number of its input morphisms) and its *codegree* (a number of its output morphisms). Similarly, any decomposable system S in **Vect** has a *reachability map* given by an epimorphism $e = A \circ B : U \rightarrow X$ and its dual *observability map* given by a monomorphism $m = C \circ A : X \rightarrow Y$; their composition $r = m \circ e : U \rightarrow Y$ in **Mor(Vect)** defines the total system's response in **Vect** given by the unique solution of the continual-sequential state equation (4.66) [Ivancevic and Snoswell (2001)].

The dual of the total system's response, defined by the feedback equation (4.67), is the *feedback morphism* $K = e^{-1} \circ m^{-1} : Y \rightarrow U$ belonging to the dual category **Vect**^{*}.

In this way, the linear, closed-loop, continual-sequential MIMO-system (4.66–4.67) represents the *linear iterative loop functor* $\mathcal{L} : \mathbf{Vect} \Rightarrow \mathbf{Vect}^*$.

Our second generalization represents a *natural system process* $\Xi[\mathcal{L}]$, that transforms the linear loop functor $\mathcal{L} : \mathbf{Vect} \Rightarrow \mathbf{Vect}^*$ – into the *nonlinear*

loop functor $\mathcal{NL} : \mathcal{CAT} \Rightarrow \mathcal{CAT}^*$ between two mutually dual nonlinear categories \mathcal{CAT} and \mathcal{CAT}^* . We apply the natural process Ξ , separately

- (1) To the feedforward decomposable system
 $S \equiv (X, A, U, B, Y, C)$ in \mathbf{Vect} , and
- (2) To the feedback morphism $K = e^{-1} \circ m^{-1} : Y \rightarrow U$ in \mathbf{Vect}^* .

Under the action of the natural process Ξ , the linear feedforward system dynamics (X, A) in \mathbf{Vect} transforms into a nonlinear feedforward Ξ --dynamics $(\Xi[X], \Xi[A])$ in \mathcal{CAT} , represented by a *nonlinear feedforward decomposable system*, $\Xi[S] \equiv (\Xi[X], \Xi[A], \Xi[U], \Xi[B], \Xi[Y], \Xi[C])$.

The reachability map transforms into the *input process* $\Xi[e] = \Xi[A] \circ \Xi[B] : \Xi[U] \rightarrow \Xi[X]$, while its dual, observability map transforms into the *output process* $\Xi[m] = \Xi[C] \circ \Xi[A] : \Xi[X] \rightarrow \Xi[Y]$. In this way the total response of the linear system $r = m \circ e : U \rightarrow Y$ in $\mathbf{Mor}(\mathbf{Vect})$ transforms into the *nonlinear system behavior*, $\Xi[r] = \Xi[m] \circ \Xi[e] : \Xi[U] \rightarrow \Xi[Y]$ in $\mathbf{Mor}(\mathcal{CAT})$. Obviously, $\Xi[r]$, if exists, is given by a nonlinear Ξ --transform of the linear state equations (4.66–4.67).

Analogously, the linear feedback morphism $K = e^{-1} \circ m^{-1} : Y \rightarrow U$ in $\mathbf{Mor}(\mathbf{Vect}^*)$ transforms into the nonlinear feedback morphism $\Xi[K] = \Xi[e^{-1}] \circ \Xi[m^{-1}] : \Xi[Y] \rightarrow \Xi[U]$ in $\mathbf{Mor}(\mathcal{CAT}^*)$.

In this way, the natural system process $\Xi : \mathcal{L} \Rightarrow \mathcal{NL}$ is established. That means that the nonlinear loop functor $L = \Xi[\mathcal{L}] : \mathcal{CAT} \Rightarrow \mathcal{CAT}^*$ is defined out of the linear, closed-loop, continual-sequential MIMO-system (4.66).

In this section we formulate the nonlinear loop functor $L = \Xi[\mathcal{L}] : \mathcal{CAT} \Rightarrow \mathcal{CAT}^*$ for various hierarchical levels of muscular-like FC.

4.9.6.2 Spinal Control Level

Our first task is to establish the nonlinear loop functor $L = \Xi[\mathcal{L}] : \mathcal{EX} \Rightarrow \mathcal{EX}^*$ on the category \mathcal{EX} of spinal FC-level.

Recall that our dissipative, driven δ --Hamiltonian biodynamical system on the configuration manifold M is, in local canonical-symplectic coordinates $q^i, p_i \in U_p$ on the momentum phase-space manifold T^*M , given by

autonomous equations

$$\dot{q}^i = \frac{\partial H_0}{\partial p_i} + \frac{\partial R}{\partial p_i}, \quad (i = 1, \dots, N) \quad (4.68)$$

$$\dot{p}_i = F_i - \frac{\partial H_0}{\partial q^i} + \frac{\partial R}{\partial q^i}, \quad (4.69)$$

$$q^i(0) = q_0^i, \quad p_i(0) = p_i^0, \quad (4.70)$$

including contravariant equation (4.68) – the velocity vector–field, and covariant equation (4.69) – the force 1–form, together with initial joint angles q_0^i and momenta p_i^0 . Here the physical Hamiltonian function $H_0 : T^*M \rightarrow \mathbb{R}$ represents the total biodynamical energy function, in local canonical coordinates $q^i, p_i \in U_p$ on T^*M given by

$$H_0(q, p) = \frac{1}{2} g^{ij} p_i p_j + V(q),$$

where $g^{ij} = g^{ij}(q, m)$ denotes the contravariant material metric tensor.

Now, the *control Hamiltonian function* $H_\gamma : T^*M \rightarrow \mathbb{R}$ of FC is in local canonical coordinates on T^*M defined by [Nijmeijer and van der Schaft (1990)]

$$H_\gamma(q, p, u) = H_0(q, p) - q^i u_i, \quad (i = 1, \dots, N) \quad (4.71)$$

where $u_i = u_i(t, q, p)$ are feedback–control 1–forms, representing the spinal FC–level u –corrections to the covariant torques $F_i = F_i(t, q, p)$.

Using δ –Hamiltonian biodynamical system (4.68–4.70) and the control Hamiltonian function (4.71), control γ_δ –Hamiltonian FC–system can be defined as

$$\begin{aligned} \dot{q}^i &= \frac{\partial H_\gamma(q, p, u)}{\partial p_i} + \frac{\partial R(q, p)}{\partial p_i}, \\ \dot{p}_i &= F_i - \frac{\partial H_\gamma(q, p, u)}{\partial q^i} + \frac{\partial R(q, p)}{\partial q^i}, \\ o^i &= -\frac{\partial H_\gamma(q, p, u)}{\partial u_i}, \quad (i = 1, \dots, N) \\ q^i(0) &= q_0^i, \quad p_i(0) = p_i^0, \end{aligned}$$

where $o^i = o^i(t)$ represent FC natural outputs which can be different from commonly used joint angles.

If nominal reference outputs $o_R^i = o_R^i(t)$ are known, the simple PD *stiffness-servo* [Whitney (1987)] could be formulated, via *error function*

$e(t) = o^j - o_R^j$, in covariant form

$$u_i = K_o \delta_{ij} (o^j - o_R^j) + K_{\dot{o}} \delta_{ij} (\dot{o}^j - \dot{o}_R^j), \quad (4.72)$$

where K s are the control-gains and δ_{ij} is the Kronecker tensor.

If natural outputs o^i *actually* are the joint angles and nominal canonical trajectories ($q_R^i = q_R^i(t)$, $p_i^R = p_i^R(t)$) are known, then the stiffness-servo (4.72) could be formulated in canonical form as

$$u_i = K_q \delta_{ij} (q^i - q_R^i) + K_p (p_i - p_i^R).$$

Now, using the fuzzified μ -Hamiltonian biodynamical system with fuzzy system numbers (i.e, imprecise segment lengths, masses and moments of inertia, joint dampings and muscular actuator parameters)

$$\dot{q}^i = \frac{\partial H_0(q, p, \sigma_\mu)}{\partial p_i} + \frac{\partial R}{\partial p_i}, \quad (4.73)$$

$$\dot{p}_i = \bar{F}_i - \frac{\partial H_0(q, p, \sigma_\mu)}{\partial q^i} + \frac{\partial R}{\partial q^i}, \quad (4.74)$$

$$q^i(0) = \bar{q}_0^i, \quad p_i(0) = \bar{p}_i^0, \quad (i = 1, \dots, N), \quad (4.75)$$

(see 3.13.4.3 above) and the control Hamiltonian function (4.71), γ_μ -Hamiltonian FC-system can be defined as

$$\begin{aligned} \dot{q}^i &= \frac{\partial H_\gamma(q, p, u, \sigma_\mu)}{\partial p_i} + \frac{\partial R(q, p)}{\partial p_i}, \\ \dot{p}_i &= \bar{F}_i - \frac{\partial H_\gamma(q, p, u, \sigma_\mu)}{\partial q^i} + \frac{\partial R(q, p)}{\partial q^i}, \\ \bar{o}^i &= -\frac{\partial H_\gamma(q, p, u, \sigma_\mu)}{\partial u_i}, \quad q^i(0) = \bar{q}_0^i, \quad p_i(0) = \bar{p}_i^0, \end{aligned}$$

where $\bar{o}^i = \bar{o}^i(t)$ represent the fuzzified natural outputs.

Finally, applying stochastic forces (diffusion fluctuations $B_{ij}[q^i(t), t]$ and discontinuous jumps in the form of ND Wiener process $W^j(t)$), i.e., using the fuzzy-stochastic $[\mu\sigma]$ -Hamiltonian biodynamical system

$$dq^i = \left(\frac{\partial H_0(q, p, \sigma_\mu)}{\partial p_i} + \frac{\partial R}{\partial p_i} \right) dt, \quad (4.76)$$

$$\begin{aligned} dp_i &= B_{ij}[q^i(t), t] dW^j(t) + \\ &\quad \left(\bar{F}_i - \frac{\partial H_0(q, p, \sigma_\mu)}{\partial q^i} + \frac{\partial R}{\partial q^i} \right) dt, \end{aligned} \quad (4.77)$$

$$q^i(0) = \bar{q}_0^i, \quad p_i(0) = \bar{p}_i^0. \quad (4.78)$$

(see 3.13.4.3 above), and the control Hamiltonian function (4.71), $\gamma_{\mu\sigma}$ -Hamiltonian FC-system can be defined as

$$\begin{aligned} dq^i &= \left(\frac{\partial H_\gamma(q, p, u, \sigma_\mu)}{\partial p_i} + \frac{\partial R(q, p)}{\partial p_i} \right) dt, \\ dp_i &= B_{ij}[q^i(t), t] dW^j(t) + \\ &\quad \left(\bar{F}_i - \frac{\partial H_\gamma(q, p, u, \sigma_\mu)}{\partial q^i} + \frac{\partial R(q, p)}{\partial q^i} \right) dt, \\ d\sigma^i &= -\frac{\partial H_\gamma(q, p, u, \sigma_\mu)}{\partial u_i} dt, \quad (i = 1, \dots, N) \\ q^i(0) &= \bar{q}_0^i, \quad p_i(0) = \bar{p}_i^0. \end{aligned}$$

If we have the case that not all of the configuration joints on the configuration manifold M are active in the specified robot task, we can introduce the coupling Hamiltonians $H^j = H^j(q, p)$, $j = 1, \dots, M \leq N$, corresponding to the system's active joints, and we come to *affine Hamiltonian function* $H_a : T^*M \rightarrow \mathbb{R}$, in local canonical coordinates on T^*M given as [Nijmeijer and van der Schaft (1990)]

$$H_a(q, p, u) = H_0(q, p) - H^j(q, p) u_j. \quad (4.79)$$

Using δ -Hamiltonian biodynamical system (4.68–4.70) and the affine Hamiltonian function (4.79), affine a_δ -Hamiltonian FC-system can be defined as

$$\dot{q}^i = \frac{\partial H_0(q, p)}{\partial p_i} - \frac{\partial H^j(q, p)}{\partial p_i} u_j + \frac{\partial R}{\partial p_i}, \quad (4.80)$$

$$\dot{p}_i = F_i - \frac{\partial H_0(q, p)}{\partial q^i} + \frac{\partial H^j(q, p)}{\partial q^i} u_j + \frac{\partial R}{\partial q^i}, \quad (4.81)$$

$$\dot{o}^i = -\frac{\partial H_a(q, p, u)}{\partial u_i} = H^j(q, p), \quad (4.82)$$

$$\begin{aligned} q^i(0) &= \bar{q}_0^i, \quad p_i(0) = \bar{p}_i^0, \\ (i &= 1, \dots, N; \quad j = 1, \dots, M \leq N). \end{aligned} \quad (4.83)$$

Using the Lie-derivative exact feedback linearization (see (4.9.2) above), and applying the *constant relative degree* r (see [Isidori (1989); Sastri and Isidori (1989)]) to all N joints of the affine a_δ -Hamiltonian FC-system (4.80–4.83), the control law for asymptotic tracking the reference outputs

σ_R^j could be formulated as

$$u_j = \frac{\dot{o}_R^{(r)j} - \mathcal{L}_f^{(r)} H^j + \sum_{s=1}^r \gamma_{s-1} (o_R^{(s-1)j} - \mathcal{L}_f^{(s-1)} H^j)}{\mathcal{L}_g \mathcal{L}_f^{(r-1)} H^j},$$

where standard MIMO-vector-fields f and g are given by

$$f = \left(\frac{\partial H_0}{\partial p_i}, -\frac{\partial H_0}{\partial q^i} \right), \quad g = \left(-\frac{\partial H^j}{\partial p_i}, \frac{\partial H^j}{\partial q^i} \right)$$

and γ_{s-1} are the coefficients of linear differential equation of order r for the error function $e(t) = o^j - \sigma_R^j$

$$e^{(r)} + \gamma_{r-1} e^{(r-1)} + \dots + \gamma_1 e^{(1)} + \gamma_0 e = 0.$$

Using the fuzzified μ -Hamiltonian biodynamical system (4.73–4.75) and the affine Hamiltonian function (4.79), affine a_μ -Hamiltonian FC-system can be defined as

$$\begin{aligned} \dot{q}^i &= \frac{\partial H_0(q, p, \sigma_\mu)}{\partial p_i} - \frac{\partial H^j(q, p, \sigma_\mu)}{\partial p_i} u_j + \frac{\partial R(q, p)}{\partial p_i}, \\ \dot{p}_i &= \bar{F}_i - \frac{\partial H_0(q, p, \sigma_\mu)}{\partial q^i} + \frac{\partial H^j(q, p, \sigma_\mu)}{\partial q^i} u_j + \frac{\partial R(q, p)}{\partial q^i}, \\ \dot{\sigma}^i &= -\frac{\partial H_a(q, p, u, \sigma_\mu)}{\partial u_i} = H^j(q, p, \sigma_\mu), \\ q^i(0) &= \bar{q}_0^i, \quad p_i(0) = \bar{p}_i^0, \quad (i = 1, \dots, N; j = 1, \dots, M \leq N). \end{aligned}$$

Using the fuzzy-stochastic $[\mu\sigma]$ -Hamiltonian biodynamical system (4.76–4.78) and the affine Hamiltonian function (4.79), affine $a_{\mu\sigma}$ -Hamiltonian FC-system can be defined as

$$\begin{aligned} dq^i &= \left(\frac{\partial H_0(q, p, \sigma_\mu)}{\partial p_i} - \frac{\partial H^j(q, p, \sigma_\mu)}{\partial p_i} u_j + \frac{\partial R(q, p)}{\partial p_i} \right) dt, \\ dp_i &= B_{ij}[q^i(t), t] dW^j(t) + \\ &\quad \left(\bar{F}_i - \frac{\partial H_0(q, p, \sigma_\mu)}{\partial q^i} + \frac{\partial H^j(q, p, \sigma_\mu)}{\partial q^i} u_j + \frac{\partial R(q, p)}{\partial q^i} \right) dt, \\ d\sigma^i &= -\frac{\partial H_a(q, p, u, \sigma_\mu)}{\partial u_i} dt = H^j(q, p, \sigma_\mu) dt, \\ q^i(0) &= \bar{q}_0^i, \quad p_i(0) = \bar{p}_i^0, \quad (i = 1, \dots, N; j = 1, \dots, M \leq N). \end{aligned}$$

Being high-degree and highly nonlinear, all of these affine control systems are extremely sensitive upon the variation of parameters, inputs,

and initial conditions. The sensitivity function S of the affine Hamiltonian $H_a(q, p, u)$ upon the parameters β_i (representing segment lengths L_i , masses m_i , moments of inertia J_i and joint dampings b_i , see [Ivancevic and Snoswell (2001); Ivancevic (1991)]), is in the case of a_δ –Hamiltonian FC–system defined as

$$S(H, \beta) = \frac{\beta_i}{H_a(q, p, u)} \frac{\partial H_a(q, p, u)}{\partial \beta_i},$$

and similarly in other two a_μ – and $a_{\mu\sigma}$ – cases.

The three affine FC–level systems a_δ , a_μ and $a_{\mu\sigma}$, resemble (in a fuzzy–stochastic–Hamiltonian form), Houk’s autogenetic motor servo of muscle spindle and Golgi tendon proprioceptors [Houk (1979)], correcting the covariant driving torques $F_i = F_i(t, q, p)$ by local ‘reflex controls’ $u_i(t, q, p)$. They form the nonlinear loop functor $L = \Xi[\mathcal{L}] : \mathcal{E}\mathcal{X} \Rightarrow \mathcal{E}\mathcal{X}^*$.

4.9.6.3 Cerebellar Control Level

Our second task is to establish the nonlinear loop functor $L = \Xi[\mathcal{L}] : \mathcal{TA} \Rightarrow \mathcal{TA}^*$ on the category \mathcal{TA} of the cerebellar FC–level. Here we propose an oscillatory neurodynamical (x, y, ω) –system (adapted from [Ivancevic *et al.* (1999a)]), a bidirectional, self–organized, associative–memory machine, resembling the function of a set of excitatory granule cells and inhibitory Purkinje cells in the middle layer of the cerebellum (see [Eccles *et al.* (1967); Houk *et al.* (1996)]). The neurodynamical (x, y, ω) –system acts on *neural–image manifold* M_{im}^N of the configuration manifold M^N as a pair of smooth, ‘1 – 1’ and ‘onto’ maps (Ψ, Ψ^{-1}) , where $\Psi : M^N \rightarrow M_{im}^N$ represents the feedforward map, and $\Psi^{-1} : M_{im}^N \rightarrow M^N$ represents the feedback map. Locally, it is defined in Riemannian neural coordinates $x^i, y_i \in V_y$ on M_{im}^N , which are in bijective correspondence with symplectic joint coordinates $q^i, p_i \in U_p$ on T^*M .

The (x, y, ω) –system is formed out of two distinct, yet nonlinearly–coupled neural subsystems, with $A^i(q)$ (4.86) and $B_i(p)$ (4.87) as system inputs, and the feedback–control 1–forms u_i (4.92) as system outputs:

- (1) Granule cells excitatory (contravariant) and Purkinje cells inhibitory (covariant) activation (x, y) –dynamics (4.84–4.87), defined respectively by a vector–field $x^i = x^i(t) : M \rightarrow TM$, representing a cross–section of the tangent bundle TM , and a 1–form $y_i = y_i(t) : M \rightarrow T^*M$, representing a cross–section of the cotangent bundle T^*M ; and

- (2) Excitatory and inhibitory unsupervised learning (ω)–dynamics (4.87–4.89) generated by random differential Hebbian learning process (4.90–4.92), defined respectively by contravariant synaptic tensor–field $\omega^{ij} = \omega^{ij}(t) : M \rightarrow TTM_{im}^N$ and covariant synaptic tensor–field $\omega_{ij} = \omega_{ij}(t) : M \rightarrow T^*T^*M$, representing cross–sections of contravariant and covariant tensor bundles, respectively.

The system equations are defined as

$$\dot{x}^i = A^i(q) + \omega^{ij} f_j(y) - x^i, \quad (4.84)$$

$$\dot{y}_i = B_i(p) + \omega_{ij} f^j(x) - y_i, \quad (4.85)$$

$$A^i(q) = K_q(q^i - q_R^i), \quad (4.86)$$

$$B_i(p) = K_p(p_i^R - p_i), \quad (4.87)$$

$$\dot{\omega}^{ij} = -\omega^{ij} + I^{ij}(x, y), \quad (4.88)$$

$$\dot{\omega}_{ij} = -\omega_{ij} + I_{ij}(x, y), \quad (4.89)$$

$$I^{ij} = f^i(x) f^j(y) + \dot{f}^i(x) \dot{f}^j(y) + \sigma^{ij}, \quad (4.90)$$

$$I_{ij} = f_i(x) f_j(y) + \dot{f}_i(x) \dot{f}_j(y) + \sigma_{ij}, \quad (4.91)$$

$$u_i = \frac{1}{2}(\delta_{ij} x^i + y_i), \quad (i, j = 1, \dots, N). \quad (4.92)$$

Here ω is a symmetric 2nd order synaptic tensor–field; $I^{ij} = I^{ij}(x, y, \sigma)$ and $I_{ij} = I_{ij}(x, y, \sigma)$ respectively denote contravariant–excitatory and covariant–inhibitory random differential Hebbian innovation–functions with tensorial Gaussian noise σ (in both variances); f s and \dot{f} s denote sigmoid activation functions ($f = \tanh(\cdot)$) and corresponding signal velocities ($\dot{f} = 1 - f^2$), respectively in both variances;

$A^i(q)$ and $B_i(p)$ are contravariant–excitatory and covariant–inhibitory neural inputs to granule and Purkinje cells, respectively; u_i are the corrections to the feedback–control 1–forms on the cerebellar FC–level.

Nonlinear activation (x, y) –dynamics (4.84–4.87), describes a two–phase biological neural oscillator field, in which excitatory neural field excites inhibitory neural field, which itself reciprocally inhibits the excitatory one. (x, y) –dynamics represents a nonlinear extension of a linear, Lyapunov–stable, conservative, gradient system, defined in local neural coordinates $x^i, y_i \in V_y$ on T^*M as

$$\dot{x}^i = -\frac{\partial \Phi}{\partial y_i} = \omega^{ij} y_j - x^i, \quad \dot{y}_i = -\frac{\partial \Phi}{\partial x^i} = \omega_{ij} x^j - y_i. \quad (4.93)$$

The gradient system (4.93) is derived from scalar, neuro-synaptic action

potential $\Phi : T^*M \rightarrow \mathbb{R}$, given by a negative, smooth bilinear form in $x^i, y_i \in V_y$ on T^*M as

$$-2\Phi = \omega_{ij}x^ix^j + \omega^{ij}y_iy_j - 2x^iy_i, \quad (i, j = 1, \dots, N), \quad (4.94)$$

which itself represents a Ψ —image of the Riemannian metrics $g : TM \rightarrow \mathbb{R}$ on the configuration manifold M .

The nonlinear oscillatory activation (x, y) —dynamics (4.84–4.87) is get from the linear conservative dynamics (4.93) by adding configuration-dependent inputs A^i and B_i , as well as sigmoid activation functions f_j and f^j , respectively. It represents an interconnected pair of excitatory and inhibitory neural fields.

Both variant-forms of learning (ω) —dynamics (4.88–4.89) are given by generalized unsupervised (self-organizing) Hebbian learning scheme (see [Kosko (1992)]) in which $\dot{\omega}_{ij}$ (resp. $\dot{\omega}^{ij}$) denotes the new-update value, $-\omega_{ij}$ (resp. $-\omega^{ij}$) corresponds to the old value and $I_{ij}(x^i, y_j)$ (resp. $I^{ij}(x^i, y_j)$) is the innovation function of the symmetric 2nd order synaptic tensor-field ω . The nonlinear innovation functions I_{ij} and I^{ij} are defined by random differential Hebbian learning process (4.90–4.91). As ω is symmetric and zero-trace coupling synaptic tensor, the conservative linear activation dynamics (4.93) is equivalent to the rule that the state of each neuron (in both neural fields) is changed in time iff the scalar action potential Φ (4.94), is lowered. Therefore, the scalar action potential Φ represents the monotonically decreasing Lyapunov function (such that $\dot{\Phi} \leq 0$) for the conservative linear dynamics (4.93), which converges to a local minimum or ground state of Φ . That is to say, the system (4.93) moves in the direction of decreasing the scalar action potential Φ , and when both $\dot{x}^i = 0$ and $\dot{y}_i = 0$ for all $i = 1, \dots, N$, the steady state is reached.

In this way, the neurodynamical (x, y, ω) —system acts as tensor-invariant self-organizing (excitatory / inhibitory) associative memory machine, resembling the set of granule and Purkinje cells of cerebellum [Houk *et al.* (1996)].

The feedforward map $\Psi : M \rightarrow M$ is realized by the inputs $A^i(q)$ and $B_i(p)$ to the (x, y, ω) —system, while the feedback map $\Psi^{-1} : M \rightarrow M$ is realized by the system output, i.e., the feedback-control 1-forms $u_i(x, y)$. These represent the cerebellar FC-level corrections to the covariant torques $F_i = F_i(t, q, p)$.

The tensor-invariant form of the oscillatory neurodynamical (x, y, ω) —system (4.84–4.92) resembles the associative action of the granule and Purkinje cells in the tuning of the limb cortico-rubro-cerebellar recurrent net-

work [Houk *et al.* (1996)], giving the cerebellar correction $u_i(x, y)$ to the covariant driving torques $F_i = F_i(t, q, p)$. In this way (x, y, ω) -system forms the nonlinear loop functor $L = \Xi[\mathcal{L}] : \mathcal{TA} \Rightarrow \mathcal{TA}^*$.

4.9.6.4 Cortical Control Level

Our third task is to establish the nonlinear loop functor $L = \Xi[\mathcal{L}] : \mathcal{ST} \Rightarrow \mathcal{ST}^*$ on the category \mathcal{ST} of the cortical FC-level.

Recall that for the purpose of cortical control, the purely rotational biodynamical manifold M could be firstly reduced to N -torus and subsequently transformed to N -cube ('hyper-joystick'), using the following geometrical techniques (see (3.8.4.3) above).

Denote by S^1 the constrained unit circle in the complex plane. This is an Abelian Lie group. We have two reduction homeomorphisms

$$SO(3) \gtrsim SO(2) \triangleright SO(2) \triangleright SO(2), \quad \text{and} \quad SO(2) \approx S^1,$$

where ' \triangleright ' denotes the noncommutative semidirect product.

Next, let I^N be the unit cube $[0, 1]^N$ in \mathbb{R}^N and ' \sim ' an equivalence relation on \mathbb{R}^N get by 'gluing' together the opposite sides of I^N , preserving their orientation. Therefore, M can be represented as the quotient space of \mathbb{R}^N by the space of the integral lattice points in \mathbb{R}^N , that is a constrained torus T^N :

$$\mathbb{R}^N / Z^N = I^N / \sim \cong \prod_{i=1}^N S_i^1 \equiv \{(q^i, i = 1, \dots, N) : \text{mod } 2\pi\} = T^N.$$

In the same way, the momentum phase-space manifold T^*M can be represented by T^*T^N .

Conversely by 'ungluing' the configuration space we get the primary unit cube. Let ' \sim^* ' denote an equivalent decomposition or 'ungluing' relation. By the *Tychonoff product-topology Theorem*, for every such quotient space there exists a 'selector' such that their quotient models are homeomorphic, that is, $T^N / \sim^* \approx A^N / \sim^*$. Therefore I_q^N represents a 'selector' for the configuration torus T^N and can be used as an N -directional ' \hat{q} -command-space' for FC. Any subset of DOF on the configuration torus T^N representing the joints included in the general biodynamics has its simple, rectangular image in the rectified \hat{q} -command space - selector I_q^N , and any joint angle q^i has its rectified image \hat{q}^i .

In the case of an end-effector, \hat{q}^i reduces to the position vector in external-Cartesian coordinates z^r ($r = 1, \dots, 3$). If orientation of the end-

effector can be neglected, this gives a topological solution to the standard inverse kinematics problem.

Analogously, all momenta \hat{p}_i have their images as rectified momenta \hat{p}_i in the \hat{p} -command space – selector I_p^N . Therefore, the total momentum phase-space manifold T^*T^N gets its ‘cortical image’ as the $(\widehat{q}, \widehat{p})$ -command space, a trivial 2ND bundle $I_q^N \times I_p^N$.

Now, the simplest way to perform the feedback FC on the cortical $(\widehat{q}, \widehat{p})$ -command space $I_q^N \times I_p^N$, and also to mimic the cortical-like behavior [1,2], is to use the 2ND fuzzy-logic controller, in pretty much the same way as in popular ‘inverted pendulum’ examples [Kosko (1992); Kosko (1996)].

We propose the fuzzy feedback-control map Ξ that maps all the rectified joint angles and momenta into the feedback-control 1-forms

$$\Xi : (\hat{q}^i(t), \hat{p}_i(t)) \mapsto u_i(t, q, p), \quad (4.95)$$

so that their corresponding universes of discourse, $\hat{M}^i = (\hat{q}_{max}^i - \hat{q}_{min}^i)$, $\hat{P}_i = (\hat{p}_i^{max} - \hat{p}_i^{min})$ and $U_i = (u_i^{max} - u_i^{min})$, respectively, are mapped as

$$\Xi : \prod_{i=1}^N \hat{M}^i \times \prod_{i=1}^N \hat{P}_i \rightarrow \prod_{i=1}^N U_i. \quad (4.96)$$

The $2N$ -D map Ξ (4.95–4.96) represents a *fuzzy inference system*, defined by (adapted from [Ivancevic *et al.* (1999b)]):

- (1) *Fuzzification* of the crisp *rectified* and *discretized* angles, momenta and controls using Gaussian-bell membership functions

$$\mu_k(\chi) = \exp\left[-\frac{(\chi - m_k)^2}{2\sigma_k}\right], \quad (k = 1, 2, \dots, 9),$$

where $\chi \in D$ is the common symbol for \hat{q}^i , \hat{p}_i and $u_i(q, p)$ and D is the common symbol for M^i , \hat{P}_i and i ; the mean values m_k of the seven partitions of each universe of discourse D are defined as $m_k = \lambda_k D + \chi_{min}$, with partition coefficients λ_k uniformly spanning the range of D , corresponding to the set of nine linguistic variables $L = \{NL, NB, NM, NS, ZE, PS, PM, PB, PL\}$; standard deviations are kept constant $\sigma_k = D/9$. Using the linguistic vector L , the 9×9 FAM (fuzzy associative memory) matrix (a ‘linguistic phase-plane’), is heuristically defined for each human joint, in a symmetrical weighted

form

$$\mu_{kl} = \varpi_{kl} \exp\{-50[\lambda_k + u(q, p)]^2\}, \quad (k, l = 1, 2, \dots, 9)$$

with weights $\varpi_{kl} \in \{0.6, 0.6, 0.7, 0.7, 0.8, 0.8, 0.9, 0.9, 1.0\}$.

- (2) *Mamdani inference* is used on each FAM-matrix μ_{kl} for all human joints:
- (i) $\mu(\hat{q}^i)$ and $\mu(\hat{p}_i)$ are combined inside the fuzzy IF-THEN rules using AND (Intersection, or Minimum) operator,

$$\mu_k[\bar{u}_i(q, p)] = \min_l \{\mu_{kl}(\hat{q}^i), \mu_{kl}(\hat{p}_i)\}.$$

(ii) the output sets from different IF-THEN rules are then combined using OR (Union, or Maximum) operator, to get the final output, fuzzy-covariant torques,

$$\mu[u_i(q, p)] = \max_k \{\mu_k[\bar{u}_i(q, p)]\}.$$

- (3) *Defuzzification* of the fuzzy controls $\mu[u_i(q, p)]$ with the ‘center of gravity’ method

$$u_i(q, p) = \frac{\int \mu[u_i(q, p)] du_i}{\int du_i},$$

to update the crisp feedback-control 1-forms $u_i = u_i(t, q, p)$. These represent the cortical FC-level corrections to the covariant torques $F_i = F_i(t, q, p)$.

Operationally, the construction of the cortical $\widehat{(q, p)}$ -command space $I_q^N \times I_p^N$ and the 2ND feedback map Ξ (4.95–4.96), mimic the regulation of locomotor conditioned reflexes by the motor cortex [Houk *et al.* (1996)], giving the cortical correction to the covariant driving torques F_i . Together they form the nonlinear loop functor $\mathcal{NL} = \Xi[\mathcal{L}] : \mathcal{ST} \Rightarrow \mathcal{ST}^*$.

A sample output from the leading human-motion simulator, *Human Biodynamics Engine* (developed by the authors in Defence Science & Technology Organisation, Australia), is given in Figure 4.9, giving the sophisticated 264 DOF analysis of adult male running with the speed of 5 m/s.

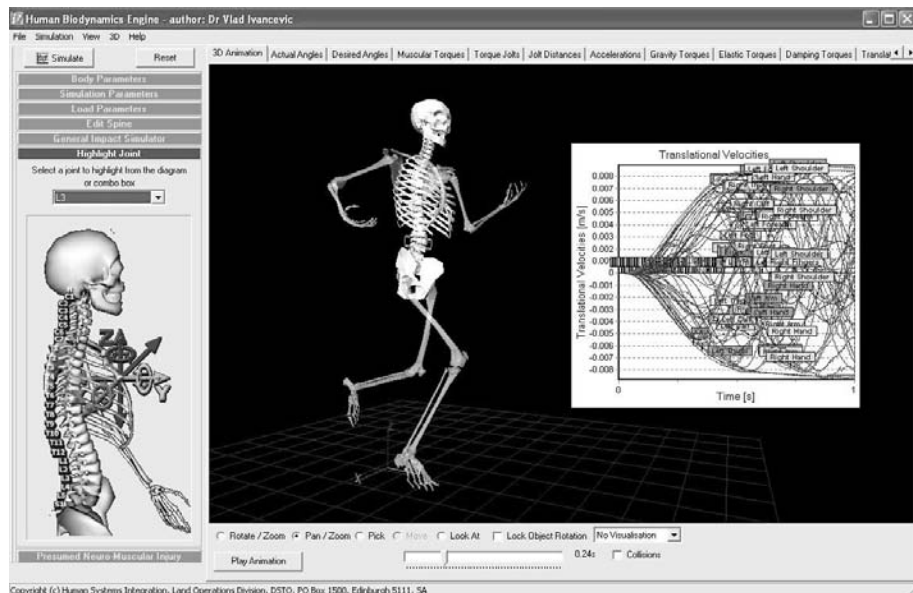


Fig. 4.9 Sample output from the Human Biodynamics Engine: running with the speed of 5 m/s.

4.9.6.5 *Open Liouville Neurodynamics and Biodynamical Self-Similarity*

Recall (see [Ivancevic and Ivancevic (2006)]) that neurodynamics has its physical behavior both at the *macroscopic*, classical, *inter-neuronal level*, and at the *microscopic*, quantum, *intra-neuronal level*. At the *macroscopic* level, various models of neural networks (NNs, for short) have been proposed as goal-oriented models of the specific neural functions, like for instance, function-approximation, pattern-recognition, classification, or control (see, e.g., [Haykin (1994)]). In the physically-based, Hopfield-type models of NNs [Hopfield (1982); Hopfield (1984)] the information is stored as a content-addressable memory in which synaptic strengths are modified after the Hebbian rule (see [Hebb (1949)]). Its retrieval is made when the network with the symmetric couplings works as the point-attractor with the fixed points. Analysis of both *activation* and *learning dynamics* of Hopfield-Hebbian NNs using the techniques of statistical mechanics [Doman *et al.* (1991)], gives us with the most important information of storage capacity, role of noise and recall performance.

On the other hand, at the general *microscopic* intra-cellular level, en-

ergy transfer across the cells, without dissipation, had been first conjectured to occur in biological matter by [Frölich and Kremer (1983)]. The phenomenon conjectured by them was based on their 1D superconductivity model: in 1D electron systems with holes, the formation of *solitonic structures* due to electron–hole pairing results in the transfer of electric current without dissipation. In a similar manner, Frölich and Kremer conjectured that energy in biological matter could be transferred without dissipation, if appropriate solitonic structures are formed inside the cells. This idea has lead theorists to construct various models for the energy transfer across the cell, based on the formation of *kink* classical solutions (see [Satarić *et al.* (1993); Satarić *et al.* (1998)]).

The interior of living cells is structurally and dynamically organized by *cytoskeletons*, i.e., networks of protein polymers. Of these structures, *microtubules* (MTs, for short) appear to be the most fundamental (see [Dustin (1984)]). Their dynamics has been studied by a number of authors in connection with the mechanism responsible for dissipation–free energy transfer. Hameroff and Penrose [Hameroff (1987)] have conjectured another fundamental role for the MTs, namely being responsible for *quantum computations* in the human neurons. [Penrose (1989); Penrose (1994); Penrose (1997)] further argued that the latter is associated with certain aspects of quantum theory that are believed to occur in the cytoskeleton MTs, in particular quantum superposition and subsequent collapse of the wave function of coherent MT networks. These ideas have been elaborated by [Mavromatos and Nanopoulos (1995a); Mavromatos and Nanopoulos (1995b)] and [Nanopoulos (1995)], based on the quantum–gravity EMN–language of [Ellis *et al.* (1992); Ellis *et al.* (1999)] where MTs have been physically modelled as non-critical (SUSY) bosonic strings. It has been suggested that the neural MTs are the microsites for the emergence of stable, macroscopic quantum coherent states, identifiable with the *preconscious states*; stringy–quantum space–time effects trigger an organized collapse of the coherent states down to a specific or *conscious state*. More recently, [Tabony *et al.* (1999)] have presented the evidence for biological self–organization and pattern formation during embryogenesis.

Now, we have two space–time biophysical scales of neurodynamics. Naturally the question arises: are these two scales somehow inter–related, is there a space–time self–similarity between them?

The purpose of this subsection is to prove the formal positive answer to the self–similarity question. We try to describe neurodynamics on both

physical levels by the *unique form* of a single equation, namely *open Liouville equation*: NN-dynamics using its classical form, and MT-dynamics using its quantum form in the Heisenberg picture. If this formulation is consistent, that would prove the *existence* of the *formal neurobiological space-time self-similarity*.

Hamiltonian Framework

Suppose that on the macroscopic NN-level we have a conservative Hamiltonian system acting in a 2ND symplectic phase-space $T^*Q = \{q^i(t), p_i(t)\}$, ($i = 1 \dots N$) (which is the cotangent bundle of the NN-configuration manifold $Q = \{q^i\}$), with a Hamiltonian function $H = H(q^i, p_i, t) : T^*Q \times \mathbb{R} \rightarrow \mathbb{R}$. The conservative dynamics is defined by classical Hamiltonian canonical equations (3.34). Recall that within the conservative Hamiltonian framework, we can apply the formalism of classical Poisson brackets: for any two functions $A = A(q^i, p_i, t)$ and $B = B(q^i, p_i, t)$ their Poisson bracket is defined as

$$[A, B] = \left(\frac{\partial A}{\partial q^i} \frac{\partial B}{\partial p_i} - \frac{\partial A}{\partial p_i} \frac{\partial B}{\partial q^i} \right).$$

Conservative Classical System

Any function $A(q^i, p_i, t)$ is called a *constant* (or *integral*) of motion of the conservative system (3.34) if

$$\dot{A} \equiv \partial_t A + [A, H] = 0, \quad \text{which implies} \quad \partial_t A = -[A, H]. \quad (4.97)$$

For example, if $A = \rho(q^i, p_i, t)$ is a *density function* of ensemble phase-points (or, a probability density to see a state $x(t) = (q^i(t), p_i(t))$ of *ensemble* at a moment t), then equation

$$\partial_t \rho = -[\rho, H] = -iL \rho \quad (4.98)$$

represents the *Liouville Theorem*, where L denotes the (Hermitian) *Liouville operator*

$$iL = [..., H] \equiv \left(\frac{\partial H}{\partial p_i} \frac{\partial}{\partial q^i} - \frac{\partial H}{\partial q^i} \frac{\partial}{\partial p_i} \right) = \text{div}(\rho \dot{\mathbf{x}}),$$

which shows that the conservative Liouville equation (4.98) is actually equivalent to the mechanical *continuity equation*

$$\partial_t \rho + \text{div}(\rho \dot{\mathbf{x}}) = 0. \quad (4.99)$$

Conservative Quantum System

We perform the formal quantization of the conservative equation (4.98) in the Heisenberg picture: all variables become Hermitian operators (denoted by ' \wedge '), the symplectic phase-space $T^*Q = \{q^i, p_i\}$ becomes the Hilbert state-space $\mathcal{H} = \mathcal{H}_{\hat{q}^i} \otimes \mathcal{H}_{\hat{p}_i}$ (where $\mathcal{H}_{\hat{q}^i} = \mathcal{H}_{\hat{q}^1} \otimes \dots \otimes \mathcal{H}_{\hat{q}^N}$ and $\mathcal{H}_{\hat{p}_i} = \mathcal{H}_{\hat{p}_1} \otimes \dots \otimes \mathcal{H}_{\hat{p}_N}$), the classical Poisson bracket $[,]$ becomes the quantum commutator $\{, \}$ multiplied by $-i/\hbar$

$$[,] \longrightarrow -i\{, \} \quad (\hbar = 1 \text{ in normal units}). \quad (4.100)$$

In this way the classical Liouville equation (4.98) becomes the *quantum Liouville equation*

$$\partial_t \hat{\rho} = i\{\hat{\rho}, \hat{H}\}, \quad (4.101)$$

where $\hat{H} = \hat{H}(\hat{q}^i, \hat{p}_i, t)$ is the Hamiltonian evolution operator, while

$$\hat{\rho} = P(a)|\Psi_a\rangle\langle\Psi_a|, \quad \text{with} \quad \text{Tr}(\hat{\rho}) = 1,$$

denotes the von Neumann *density matrix operator*, where each quantum state $|\Psi_a\rangle$ occurs with probability $P(a)$; $\hat{\rho} = \hat{\rho}(\hat{q}^i, \hat{p}_i, t)$ is closely related to another von Neumann concept: *entropy* $S = -\text{Tr}(\hat{\rho}[\ln \hat{\rho}])$.

Open Classical System

We now move to the *open (nonconservative) system*: on the macroscopic NN-level the *opening operation* equals to the *adding* of a *covariant* vector of external (dissipative and/or motor) forces $F_i = F_i(q^i, p_i, t)$ to (the r.h.s of) the covariant Hamiltonian *force equation*, so that Hamiltonian equations get the *open (dissipative and/or forced) form*

$$\dot{q}^i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = F_i - \frac{\partial H}{\partial q^i}. \quad (4.102)$$

In the framework of the open Hamiltonian system (4.102), dynamics of any function $A(q^i, p_i, t)$ is defined by the *open evolution equation*:

$$\partial_t A = -[A, H] + \Phi,$$

where $\Phi = \Phi(F_i)$ represents the general form of the scalar force term.

In particular, if $A = \rho(q^i, p_i, t)$ represents the density function of ensemble phase-points, then its dynamics is given by the (dissipative/forced)

open Liouville equation:

$$\partial_t \rho = -[\rho, H] + \Phi. \quad (4.103)$$

In particular, the scalar force term can be cast as a linear Poisson–bracket form

$$\Phi = F_i[A, q^i], \quad \text{with} \quad [A, q^i] = -\frac{\partial A}{\partial p_i}. \quad (4.104)$$

Now, in a similar way as the conservative Liouville equation (4.98) resembles the continuity equation (4.99) from continuum dynamics, also the open Liouville equation (4.103) resembles the probabilistic *Fokker–Planck equation* from statistical mechanics. If we have a ND stochastic process $x(t) = (q^i(t), p_i(t))$ defined by the vector *Itô SDE*

$$dx(t) = f(x, t) dt + G(x, t) dW,$$

where f is a ND vector function, W is a KD Wiener process, and G is a $N \times KD$ matrix valued function, then the corresponding probability density function $\rho = \rho(x, t | \hat{x}, t')$ is defined by the ND Fokker–Planck equation (see, e.g., [Gardiner (1985)])

$$\partial_t \rho = -\text{div}[\rho f(x, t)] + \frac{1}{2} \frac{\partial^2}{\partial x_i \partial x_j} (Q_{ij} \rho), \quad (4.105)$$

where $Q_{ij} = (G(x, t) G^T(x, t))_{ij}$. It is obvious that the Fokker–Planck equation (4.105) represents the particular, stochastic form of our general open Liouville equation (4.103), in which the scalar force term is given by the (second–derivative) noise term

$$\Phi = \frac{1}{2} \frac{\partial^2}{\partial x_i \partial x_j} (Q_{ij} \rho).$$

Equation (4.103) will represent the open classical model of our macroscopic NN–dynamics.

Continuous Neural Network Dynamics

The generalized NN–dynamics, including two special cases of *graded response neurons* (GRN) and *coupled neural oscillators* (CNO), can be presented in the form of a stochastic *Langevin rate equation*

$$\dot{\sigma}_i = f_i + \eta_i(t), \quad (4.106)$$

where $\sigma_i = \sigma_i(t)$ are the continual neuronal variables of i th neurons (representing either membrane action potentials in case of GRN, or oscillator phases in case of CNO); J_{ij} are individual synaptic weights; $f_i = f_i(\sigma_i, J_{ij})$ are the deterministic forces (given, in GRN-case, by $f_i = \sum_j J_{ij} \tanh[\gamma \sigma_j] - \sigma_i + \theta_i$, with $\gamma > 0$ and with the θ_i representing injected currents, and in CNO-case, by $f_i = \sum_j J_{ij} \sin(\sigma_j - \sigma_i) + \omega_i$, with ω_i representing the natural frequencies of the individual oscillators); the noise variables are given as $\eta_i(t) = \lim_{\Delta \rightarrow 0} \zeta_i(t) \sqrt{2T/\Delta}$ where $\zeta_i(t)$ denote uncorrelated Gaussian distributed random forces and the parameter T controls the amount of noise in the system, ranging from $T = 0$ (deterministic dynamics) to $T = \infty$ (completely random dynamics).

More convenient description of the neural random process (4.106) is provided by the Fokker–Planck equation describing the time evolution of the probability density $P(\sigma_i)$

$$\partial_t P(\sigma_i) = -\frac{\partial}{\partial \sigma_i} (f_i P(\sigma_i)) + T \frac{\partial^2}{\partial \sigma_i^2} P(\sigma_i). \quad (4.107)$$

Now, in the case of deterministic dynamics $T = 0$, equation (4.107) can be put into the form of the conservative Liouville equation (4.98), by making the substitutions: $P(\sigma_i) \rightarrow \rho$, $f_i = \dot{\sigma}_i$, and $[\rho, H] = \text{div}(\rho \dot{\sigma}_i) \equiv \sum_i \frac{\partial}{\partial \sigma_i} (\rho \dot{\sigma}_i)$, where $H = H(\sigma_i, J_{ij})$. Further, we can formally identify the stochastic forces, i.e., the second-order noise-term $T \sum_i \frac{\partial^2}{\partial \sigma_i^2} \rho$ with $F^i[\rho, \sigma_i]$, to get the open Liouville equation (4.103).

Therefore, on the NN-level deterministic dynamics corresponds to the conservative system (4.98). Inclusion of stochastic forces corresponds to the system opening (4.103), implying the *macroscopic arrow of time*.

Open Quantum System

By formal quantization of equation (4.103) with the scalar force term defined by (4.104), in the same way as in the case of the conservative dynamics, we get the *quantum open Liouville equation*

$$\partial_t \hat{\rho} = i\{\hat{\rho}, \hat{H}\} + \hat{\Phi}, \quad \text{with} \quad \hat{\Phi} = -i\hat{F}_i\{\hat{\rho}, \hat{q}^i\}, \quad (4.108)$$

where $\hat{F}_i = \hat{F}_i(\hat{q}^i, \hat{p}_i, t)$ represents the covariant quantum operator of external friction forces in the Hilbert state-space $\mathcal{H} = \mathcal{H}_{\hat{q}^i} \otimes \mathcal{H}_{\hat{p}_i}$.

Equation (4.108) will represent the open quantum-friction model of our microscopic MT-dynamics. Its system-independent properties are [Ellis

et al. (1992); Ellis *et al.* (1999); Mavromatos and Nanopoulos (1995a); Mavromatos and Nanopoulos (1995b); Nanopoulos (1995)]:

(1) Conservation of probability P

$$\partial_t P = \partial_t [\text{Tr}(\hat{\rho})] = 0.$$

(2) Conservation of energy E , on the average

$$\partial_t \langle \langle E \rangle \rangle \equiv \partial_t [\text{Tr}(\hat{\rho} E)] = 0.$$

(3) Monotonic increase in entropy

$$\partial_t S = \partial_t [-\text{Tr}(\hat{\rho} \ln \hat{\rho})] \geq 0,$$

and thus automatically and naturally implies a *microscopic arrow of time*, so essential in realistic biophysics of neural processes.

Non-Critical Stringy MT-Dynamics

In EMN-language of non-critical (SUSY) bosonic strings, our MT-dynamics equation (4.108) reads

$$\partial_t \hat{\rho} = i\{\hat{\rho}, \hat{H}\} - i\hat{g}_{ij}\{\hat{\rho}, \hat{q}^i\}\hat{q}^j, \quad (4.109)$$

where the target-space density matrix $\hat{\rho}(\hat{q}^i, \hat{p}_i)$ is viewed as a function of coordinates \hat{q}^i that parameterize the couplings of the generalized σ -models on the bosonic string world-sheet, and their conjugate momenta \hat{p}_i , while $\hat{g}_{ij} = \hat{g}_{ij}(\hat{q}^i)$ is the quantum operator of the *positive definite metric* in the space of couplings. Therefore, the covariant quantum operator of external friction forces is in EMN-formulation given as $\hat{F}_i(\hat{q}^i, \hat{q}^i) = \hat{g}_{ij} \hat{q}^j$.

Equation (4.109) establishes the conditions under which a large-scale coherent state appearing in the MT-network, which can be considered responsible for loss-free energy transfer along the tubulins.

Equivalence of Neurodynamic Forms

It is obvious that both the macroscopic NN-equation (4.103) and the microscopic MT-equation (4.108) have the same open Liouville form, which implies the arrow of time. These proves the existence of the formal neuro-biological space-time self-similarity.

In this way, we have described neurodynamics of both NN and MT ensembles, belonging to completely different biophysical space-time scales,

by the unique form of open Liouville equation, which implies the arrow of time. The existence of the formal *neuro-biological self-similarity* has been proved.

4.9.7 Brain–Mind Functorial Machines

In this section we propose two models of the *brain–mind functorial machines*: the first one is a psychologically–motivated top–down machine, while the second one is physically–motivated bottom–up solitary machine.

4.9.7.1 Neurodynamical 2–Functor

Here we define the goal–directed cognitive neurodynamics as an evolution 2–functor \mathcal{E} given by

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ \downarrow h & \text{CURRENT NEURAL STATE} & \downarrow g \\ C & \xrightarrow{k} & D \end{array}$$

$\xRightarrow{\mathcal{E}}$

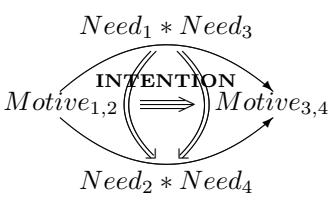
$$\begin{array}{ccc} \mathcal{E}(A) & \xrightarrow{\mathcal{E}(f)} & \mathcal{E}(B) \\ \downarrow \mathcal{E}(h) & \text{DESIRED NEURAL STATE} & \downarrow \mathcal{E}(g) \\ \mathcal{E}(C) & \xrightarrow{\mathcal{E}(k)} & \mathcal{E}(D) \end{array}$$

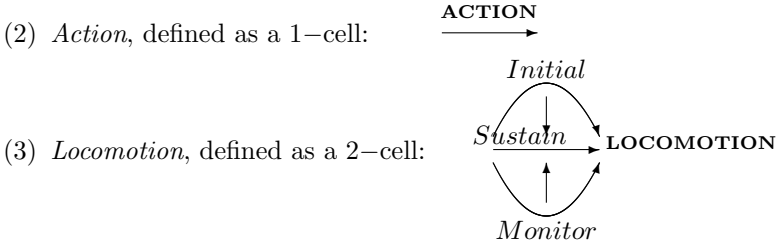
(4.110)

In (4.110), \mathcal{E} represents a projection map from the *source 2–category* of the current neural state, defined as a commutative square of small categories A, B, C, D, \dots of *current neural ensembles* and their causal interrelations f, g, h, k, \dots , onto the *target 2–category* of the desired neural state, defined as a commutative square of small categories $\mathcal{E}(A), \mathcal{E}(B), \mathcal{E}(C), \mathcal{E}(D), \dots$ of *evolved neural ensembles* and their causal interrelations $\mathcal{E}(f), \mathcal{E}(g), \mathcal{E}(h), \mathcal{E}(k)$.

The evolution 2–functor \mathcal{E} can be horizontally decomposed in the following three neurodynamic components (see [Lewin (1997); Aidman and Leontiev (1991)]):

- (1) *Intention*, defined as a 3–cell:





Now, each causal arrow in (4.110), say $f : A \rightarrow B$, stands for a generic ‘neuro-morphism’, representing a self-organized, oscillatory neurodynamic system. We define a generic neuro-morphism f to be a nonlinear tensor-field (x, y, ω) -system (4.111–4.116), acting as a bidirectional associative memory machine on a ND Riemannian manifold M^N of the human cortex. It is formed out of two distinct, yet nonlinearly-coupled neural subsystems:

- (1) Activation (x, y) -dynamics (4.111–4.112), defined as an interplay of an excitatory vector-field $x^i = x^i(t) : M^N \rightarrow TM$, representing a cross-section of the tangent bundle TM , and an inhibitory 1-form $y_i = y_i(t) : M^N \rightarrow T^*M$, representing a cross-section of the cotangent bundle T^*M .
- (2) Excitatory and inhibitory unsupervised learning (ω) -dynamics (4.113–4.116) generated by random differential Hebbian learning process (4.115–4.116), defined respectively by contravariant synaptic tensor-field $\omega^{ij} = \omega^{ij}(t) : M^N \rightarrow TTM_{im}^N$ and covariant synaptic tensor-field $\omega_{ij} = \omega_{ij}(t) : M^N \rightarrow T^*T^*M$, representing cross-sections of contravariant and covariant tensor bundles, respectively.

(x, y, ω) -system is analytically defined as a set of N coupled neurodynamic equations:

$$\dot{x}^i = A^i + \omega^{ij} f_j(y) - x^i, \quad (4.111)$$

$$\dot{y}_i = B_i + \omega_{ij} f^j(x) - y_i, \quad (4.112)$$

$$\dot{\omega}^{ij} = -\omega^{ij} + I^{ij}(x, y), \quad (4.113)$$

$$\dot{\omega}_{ij} = -\omega_{ij} + I_{ij}(x, y), \quad (4.114)$$

$$I^{ij} = f^i(x) f^j(y) + \dot{f}^i(x) \dot{f}^j(y) + \sigma^{ij}, \quad (4.115)$$

$$I_{ij} = f_i(x) f_j(y) + \dot{f}_i(x) \dot{f}_j(y) + \sigma_{ij}, \quad (4.116)$$

$$(i, j = 1, \dots, N).$$

Here ω is a symmetric, second-order synaptic tensor-field; $I^{ij} = I^{ij}(x, y, \sigma)$ and $I_{ij} = I_{ij}(x, y, \sigma)$ respectively denote contravariant-excitatory and

covariant–inhibitory random differential Hebbian innovation–functions with tensorial Gaussian noise σ (in both variances); f s and \dot{f} s denote sigmoid activation functions ($f = \tanh(\cdot)$) and corresponding signal velocities ($\dot{f} = 1 - f^2$), respectively in both variances; $A^i = A^i(t)$ and $B_i = B_i(t)$ are contravariant–excitatory and covariant–inhibitory neural inputs to the corresponding cortical cells, respectively;

Nonlinear activation (x, y) –dynamics, describes a two–phase biological neural oscillator field, in which the excitatory neural field excites the inhibitory neural field, which itself reciprocally inhibits the excitatory one. (x, y) –dynamics represents a nonlinear extension of a linear, Lyapunov–stable, conservative, gradient system, defined in local neural coordinates $x^i, y_i \in V_y$ on T^*M as

$$\dot{x}^i = -\frac{\partial \Phi}{\partial y_i} = \omega^{ij} y_j - x^i, \quad \dot{y}_i = -\frac{\partial \Phi}{\partial x^i} = \omega_{ij} x^j - y_i. \quad (4.117)$$

The gradient system (4.117) is derived from scalar, neuro–synaptic action potential $\Phi : T^*M \rightarrow \mathbb{R}$, given by a negative, smooth bilinear form in $x^i, y_i \in V_y$ on T^*M as

$$-2\Phi = \omega_{ij} x^i x^j + \omega^{ij} y_i y_j - 2x^i y_i, \quad (i, j = 1, \dots, N),$$

which itself represents a Ψ –image of the Riemannian metrics $g : TM \rightarrow \mathbb{R}$ on the configuration manifold M^N .

The nonlinear oscillatory activation (x, y) –dynamics (4.111–4.114) is get from the linear conservative dynamics (4.117), by adding configuration dependent inputs A^i and B_i , as well as sigmoid activation functions f_j and \dot{f}^j , respectively. It represents an interconnected pair of excitatory and inhibitory neural fields.

Both variant–forms of learning (ω) –dynamics (4.113–4.114) are given by a generalized unsupervised (self–organizing) Hebbian learning scheme (see [Kosko (1992)]) in which $\dot{\omega}_{ij}$ (resp. $\dot{\omega}^{ij}$) denotes the new–update value, $-\omega_{ij}$ (resp. ω^{ij}) corresponds to the old value and $I_{ij}(x^i, y_j)$ (resp. $I^{ij}(x^i, y_j)$) is the innovation function of the symmetric 2nd order synaptic tensor–field ω . The nonlinear innovation functions I_{ij} and I^{ij} are defined by random differential Hebbian learning process (4.115–4.116). As ω is a symmetric and zero–trace coupling synaptic tensor, the conservative linear activation dynamics (4.117) is equivalent to the rule that ‘the state of each neuron (in both neural fields) is changed in time if, and only if, the scalar action potential Φ (52), is lowered’. Therefore, the scalar action potential Φ represents the monotonically decreasing Lyapunov function (such that

$\dot{\Phi} \leq 0$) for the conservative linear dynamics (4.117), which converges to a local minimum or ground state of Φ . That is to say, the system (4.117) moves in the direction of decreasing the scalar action potential Φ , and when both $\dot{x}^i = 0$ and $\dot{y}_i = 0$ for all $i = 1, \dots, N$, the steady state is reached.

4.9.7.2 Solitary ‘Thought Nets’ and ‘Emerging Mind’

Synergetic ‘Thought Solitons’

Recall that synergetics teaches us that *order parameters* (and their spatio-temporal evolution) are *patterns, emerging from chaos*. In our opinion, the most important of these order parameters, both natural and man made, are *solitons*, because of their self-organizing quality to create order out of chaos. From this perspective, *nonlinearity* – the essential characteristic of nature – is the *cause* of both *chaos* and *order*. Recall that the solitary particle-waves, also called the ‘light bullets’, are localized space-time excitations $\Psi(x, t)$, propagating through a certain medium Ω with constant velocities v_j . They describe a variety of nonlinear wave phenomena in one dimension and playing important roles in optical fibers, many branches of physics, chemistry and biology.

To derive our solitary network we start with modelling the conservative ‘*thought solitons*’, using the following three classical nonlinear equations, defining the time evolution of the spatio-temporal wave function $\Psi(x, t)$ (which is smooth, and either complex-, or real-valued) (see [Novikov *et al.* (1984); Fordy (1990); Ablowitz and Clarkson (1991); Ivancevic and Pearce (2001a)]; also compare with (3.13.2) above):

(1) Nonlinear Schrödinger (NS) equation

$$i\Psi_t = 2\mu|\Psi|^2\Psi - \Psi_{xx}, \quad (4.118)$$

where $\Psi = \Psi(x, t)$ is a complex-valued wave function with initial condition $\Psi(x, t)|_{t=0} = \Psi(x)$ and μ is a nonlinear parameter representing field strength. In the linear limit ($\mu = 0$) NS becomes the ordinary Schrödinger equation for the wave function of the free 1D particle with mass $m = 1/2$. Its Hamiltonian function

$$H_{NS} = \int_{-\infty}^{+\infty} \left(\mu|\Psi|^4 + |\Psi_x|^2 \right) dx,$$

is equal to the total and conserved energy of the soliton. NS describes,

for example, nonlinear Faraday resonance in a vertically oscillating water, an easy-plane ferromagnet with a combination of a stationary and a high-frequency magnetic fields, and the effect of phase-sensitive amplifiers on solitons propagating in optical fibers.

(2) *Korteweg-de Vries (KdV) equation*

$$\Psi_t = 6\Psi\Psi_x - \Psi_{xxx},$$

with Hamiltonian (total conserved energy) given by

$$H_{KdV} = \int_{-\infty}^{+\infty} \left(\Psi^3 + \frac{1}{2} \Psi_x^2 \right) dx.$$

KdV is related to the ordinary Schrödinger equation by the inverse scattering method. KdV is a well-known model of 1D turbulence that was derived in various physical contexts, including chemical-reaction waves, propagation of combustion fronts in gases, surface waves in a film of a viscous liquid flowing along an inclined plane, patterns in thermal convection, rapid solidification, and others. Its discretization gives the Lotka-Volterra equation

$$\dot{x}^j(t) = x^j(t) (x^{j+1}(t) - x^{j-1}(t)),$$

which appears in a model of struggle for existence of biological species.

(3) *Sine-Gordon (SG) equation*

$$\Psi_{tt} = \Psi_{xx} - \sin \Psi,$$

with Hamiltonian (total conserved energy) given by

$$H_{SG} = \int_{-\infty}^{+\infty} (\Psi_t^2 + \Psi_x^2 + \cos \Psi) dx.$$

SG gives one of the simplest models of the unified field theory, can be found in the theory of dislocations in metals, in the theory of Josephson junctions and so on. It can be used also in interpreting certain biological processes like DNA dynamics. Its discretization gives a *system of coupled pendulums*.

Discrete solitons exist also in the form of the *soliton celular automata* (SCA) [Park *et al.* (1986)]. SCA is a 1(space)+1(time)-dimensional ‘box and ball system’ made of infinite number of zeros (or, boxes) and finite

number of ones (or, balls). The value of the j th SCA cell a_t^j at a discrete time t , is given as

$$a_{t+1}^j = \begin{cases} 1, & \text{if } a_t^j = 0 \text{ and } \sum_{i=-\infty}^{j-1} u_t^i > \sum_{i=-\infty}^{j-1} a_{t+1}^i, \\ 0, & \text{otherwise,} \end{cases}$$

where $a_t^j = 0$ is assumed for $|j| \gg 1$. Any state of the SCA consists purely of solitons (particularly, KdV-solitons), possessing conserved quantities of the form of H_{KdV} . All of these properties have motivated a number of suggestive applications for a new kind of computational architecture that will use these evolution patterns of SCA in order to give a ‘gateless’ implementation of logical operations.

In practice, both SCA and KdV are usually approximated by the *Toda lattice equation*,

$$\ddot{q}^i = e^{q^{i+1}-q^i} - e^{q^i-q^{i-1}}, \quad (i = 1, \dots, N) \quad (4.119)$$

with quasiperiodic $q^{N+i}(t) = q^i(t) + c$, or,

fast-decaying boundary conditions $\lim_{i \rightarrow -\infty} q^i(t) = 0$, $\lim_{i \rightarrow +\infty} q^i(t) = c$.

The Toda equation (4.119) is a *gradient Newtonian equation* of motion

$$\ddot{q}^i = -\partial_q^i V, \quad V(q) = \sum_{i=1}^N e^{q^{i+1}-q^i}.$$

Otherwise, the Toda equation represents a Hamiltonian system

$$\dot{q}^i = p_i, \quad \dot{p}_i = e^{q^{i+1}-q^i} - e^{q^i-q^{i-1}},$$

with the phase-space $P = \mathbb{R}^{2N}$ with coordinates (p_i, q^i) , standard Poisson structure

$$\{p_i, p_j\} = \{q^i, q^j\} = 0, \quad \{p_i, q^j\} = \delta_i^j,$$

and Hamiltonian function $H = \sum_{i=1}^N \left(\frac{1}{2} p_i^2 + e^{q^{i+1}-q^i} \right)$, $(i, j = 1, \dots, N)$.

Next, to make our conservative thought solitons *open to the environment*, we have to modify them by adding:

1. Input from the senses, in the form of the Weber–Fechner’s law,

$$S(t) = a_r \log s^r(t), \quad (r = 1, \dots, 5), \quad (4.120)$$

where $S = S(t)$ is the *sensation*, $s^r = s^r(t)$ the vector of stimuli from the five senses, and a_r a constant vector; and

2. Disturbances, in the form of additive, zero-mean Gaussian white noise $\eta = \eta(t)$, independent from the main soliton-signal.

In this way, we get the *modified solitary equations*:

$$\text{MNS : } \quad i\Psi_t = 2\mu|\Psi|^2\Psi - \Psi_{xx} + a_r \log s^r \Psi + \eta,$$

$$\text{MKdV : } \quad \Psi_t = 6\Psi\Psi_x - \Psi_{xxx} + a_r \log s^r \Psi + \eta,$$

$$\text{MSG : } \quad \Psi_{tt} = \Psi_{xx} - \sin \Psi + a_r \log s^r \Psi + \eta,$$

representing the *three different models of the thought units*.

Now we will form a single emerging order-parameter, the general factor, that we call the *Mind*. It behaves like an orchestrated ensemble of thought solitons, defined as systems of trainable, coupled soliton equations. Their tensor couplings perform *self-organizing associative learning by trial and error*, similar to that of the neural ensemble.

The dynamics of the *soliton ensemble*, representing our model of the ‘mind’ can be described as one of the following three soliton systems; each of them performs *learning*, *growing* and *competing* between each other, and *communicates with environment* through the *sensory inputs* and the *heating noise*:

1. Coupled modified nonlinear Schrödinger equations

$$\begin{aligned} i\Psi_t^k = & -\Psi_{xx}^k + 2\mu^k \sum_{j \neq k} |\Psi^j|^2 \mathbf{W}_k^j S^j(\Psi^j) \\ & + \nu^k \Psi^k (1 - \epsilon^k \Psi^k) + a_r \log s^r \Psi^k + \eta^k, \end{aligned}$$

2. Coupled modified Korteweg-de Vries equations

$$\begin{aligned} \Psi_t^k = & 6\Psi_x^k \Psi^k - \Psi_{xxx}^k + \sum_{j \neq k} \mathbf{W}_k^j S^j(\Psi^j) \\ & + \nu^k \Psi^k (1 - \epsilon^k \Psi^k) + a_r \log s^r \Psi^k + \eta^k, \end{aligned}$$

3. Coupled modified Sine-Gordon equations

$$\Psi_{tt}^k = \Psi_{xx}^k - \sin \Psi^k + \sum_{j \neq k} + \nu^k \Psi^k (1 - \epsilon^k \Psi^k) + a_r \log s^r \Psi^k + \eta^k,$$

where $\Psi^k = \Psi^k(x, t)$, ($k = 1, \dots, n$) is the set of wave functions of the solitary thoughts, $S(\cdot)$ represents the sigmoidal threshold functions, ν^k and ϵ^k are *growing* and *competition* parameters.

$\mathbf{W}_k^j = \mathbf{W}_k^j(\Psi)$ are *tensorial learning couplings*, evolving according to the Hebbian learning scheme (see [Kosko (1992)]):

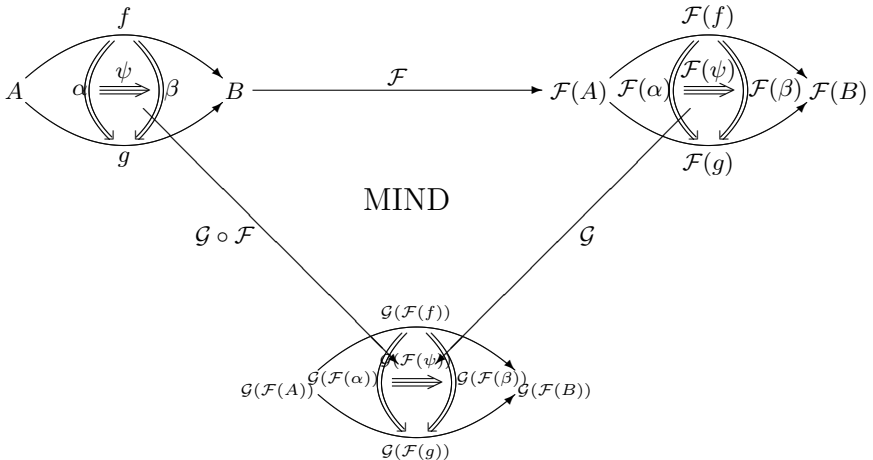
$$\dot{\mathbf{W}}_k^j = -\mathbf{W}_k^j + \Phi_k^j(\Psi^k, \Psi^j),$$

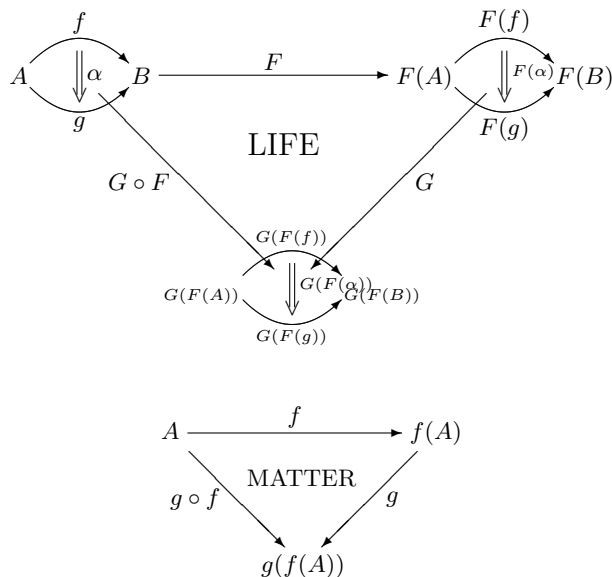
with *innovation* defined in tensor signal form (here $\dot{S}(\cdot) = 1 - \tanh(\cdot)$)

$$\Phi_k^j = S^j(\Psi^j) S^k(\Psi^k) + \dot{S}^j(\Psi^j) \dot{S}^k(\Psi^k).$$

Emerging Categorical Structure: $MATTER \Rightarrow LIFE \Rightarrow MIND$

The solitary thought nets effectively simulate the following 3–categorical structure of MIND, emerging from the 2–categorical structure of LIFE, which is itself emerging from the 1–categorical structure of MATTER:





4.9.8 Geometrodynamics of Human Crowd

In this subsection we formulate crowd representation model as an emotion–field induced collective behavior of individual autonomous agents [Ivancevic and Snoswell (2000)].

It is well-known that *crowd behavior* is more influenced by *collective emotion* than by cognition. Recall from previous subsection that according to *Lewinian psychodynamics*, human behavior is largely determined by underlying *forces* (needs). For him, a force–field is defined as ‘the totality of coexisting motivational facts which are conceived of as mutually interdependent’ [Lewin (1997)]. He also stresses psychological *direction* and *velocity* of behavior.

On the other hand, a number of factor–analysis based studies show that human emotion is not a single quantity, but rather a *multidimensional space*. For example, in [Skiffington (1998)], authors assessed emotions with single adjective descriptors using standard linear factor analysis, by examining semantic as well as cognitive, motivational, and intensity features of emotions. The focus was on seven negative emotions common to several emotion typologies: anger, fear, sadness, shame, pity, jealousy, and contempt. For each of these emotions, seven items were generated correspond-

ing to cognitive appraisal about the self, cognitive appraisal about the environment, action tendency, action fantasy, synonym, antonym, and intensity range of the emotion, respectively. These findings set the groundwork for the construction of an instrument to assess emotions multicomponentially.

4.9.8.1 Crowd Hypothesis

We consider a *human crowd* \mathcal{C} as a group of m *autonomous agents* \mathcal{A}_i ($i = 1, \dots, m$), each of which carries its own nD *motivational factor-structure*. This nonlinear factor structure, which can be get using modern *nonlinear factor analysis* techniques (see [Yalcin and Amemiya (2001); Amemiya (1993); Wall and Amemiya (1998); Wall and Amemiya (2000)]); also compare with subsection 3.11.4 below), is defined by n hypothetical motivational factor-coordinates $\mathbf{q}_i = \{q_i^\mu\}$, ($\mu = 1, \dots, n$), spanning the smooth nD *motivational factor manifold* M_i for each autonomous agent \mathcal{A}_i .

We understand crowd representation as an environmental field-induced collective behavior of individual autonomous agents. To model it in a general geometrodynamical framework, we firstly define the *behavior* of each agent \mathcal{A}_i as a *motion* π_i along his motivational manifold M_i , caused by his own *emotion field* Φ_i , which is an active (motor) subset of M_i .

Secondly, we formulate a collective geometrodynamical model for the crowd, considered as a union $\mathcal{C} = \cup_i \mathcal{A}_i$, in the form of a divergence equation for the total crowd's SEM-tensor \mathbf{C} .¹⁴

4.9.8.2 Geometrodynamics of Individual Agents

To formulate individual agents' geometrodynamics, we firstly derive two higher geometrical structures from a motivational factor manifold M_i corresponding to an agent \mathcal{A}_i : (i) the *agent's velocity phase-space*, defined as a tangent bundle TM_i , and (ii) the *agent's momentum phase-space*, defined as a cotangent bundle T^*M_i .

Now, the sections of TM_i we call the agent's vector-fields \mathbf{v}_i , which can be expanded in terms of the basis vector-fields $\{e_\mu^i \equiv \partial_{q_i^\mu}\}$, as $\mathbf{v}_i = v_i^\mu e_\mu^i$. Similarly, the sections of T^*M_i we call the agent's one-forms α_i , which can be expanded in terms of the basis one-forms $\{\omega_i^\mu \equiv dq_i^\mu\}$, as $\alpha_i = \alpha_i^\mu \omega_i^\mu$. Here d denotes the *exterior derivative* (such that $dd = 0$). In particular,

¹⁴Throughout the text we use the following index convention: we label individual agents using Latin indices, and individual motivational factors using Greek indices; summation convention is applied only to Greek factor indices.

the *velocity* vector-fields $\dot{\mathbf{q}}_i$ are defined on TM_i as $\dot{\mathbf{q}}_i = \dot{q}_i^\mu e_\mu^i$, while the *momentum* one-forms π_i are defined on T^*M_i as $\pi_i = \pi_\mu^i \omega_i^\mu$.

Also, all factor-configuration manifolds M_i are assumed to be Riemannian, admitting the metrics $\mathbf{g}_i = \langle \omega_i^\mu, \omega_i^\eta \rangle$, determined by *kinetic energies* $T_i = \langle \dot{q}_i^\mu, \dot{q}_i^\eta \rangle$ of individual agents \mathcal{A}_i , each with the metric tensor $\mathbf{g}^i = g_{\mu\eta}^i \omega_i^\mu \otimes \omega_i^\eta$. This implies that all vector-fields \mathbf{v}_i and one-forms α_i are related by \mathbf{g} -induced scalar products $\langle \omega_i^\mu, e_\eta^i \rangle = \delta_\eta^{\mu i}$.

Emotional/Environmental Fields and Induced Agents' Motions

Now, for each autonomous agent \mathcal{A}_i three additional geometrodynamical objects are defined as:

- (1) *Emotional/environmental potential* one-form $\alpha_i = \alpha_\mu^i \omega_i^\mu$, which is the gradient of some scalar function $\mathbf{f}_i = f_i^\mu(q_i^\mu)$ on M_i ,

$$\alpha_i = d\mathbf{f}_i, \quad \text{in components,} \quad \alpha_\mu^i = \partial_{q_i^\mu} f_i^\mu \omega_i^\mu.$$

- (2) *Psycho-physical current* vector-field $\mathbf{J}_i = J_i^\mu e_\mu^i$ on M_i , defined through its *motivational charge* e_i as

$$J_i^\mu = e_i \int_{t_i} \dot{q}_i^\mu \delta^n [q_i^\mu(t_i)] dt_i,$$

where $\delta^n = \delta^n [q_i^\mu(t_i)]$ denotes the n D impulse delta-function defined on M_i .

- (3) *Emotional/environmental psycho-physical field* is a two-form

$$\Phi_i = \Phi_{\mu\eta}^i \omega_i^\mu \otimes \omega_i^\eta = \frac{1}{2} \Phi_{\mu\eta}^i \omega_i^\mu \wedge \omega_i^\eta$$

on M_i , defined as the exterior derivative (i.e., curl) of the emotional/enviro-nmental potential α_i ,

$$\Phi_i = d\alpha_i, \quad \text{in components,} \quad \Phi_{\mu\eta}^i = \alpha_{\eta;\mu}^i - \alpha_{\mu;\eta}^i.$$

The emotional/environmental psycho-physical field Φ_i is governed by two standard field equations:

$$d\Phi_i = dd\alpha_i = 0, \quad \text{in components,} \quad \Phi_{[\mu\eta;\nu]}^i = 0, \quad (4.121)$$

and

$$\text{div } \Phi_i = \mathbf{g}^i \mathbf{J}_i, \quad \text{in components,} \quad \Phi_{\mu\eta}^{i;\eta} = g_{\mu\eta}^i J_i^\eta, \quad (4.122)$$

where $[\mu\eta; \nu] \equiv \mu\eta; \nu + \eta\nu; \mu + \nu\mu; \eta$. The first field equation (4.121) states that for each agent \mathcal{A}_i the motivational tension-field Φ_i is *curl-free*, while the second field equation (4.122) states that the environmental psycho-physical field Φ_i has its *source* in the psycho-physical current \mathbf{J}_i .

Now, the *behavioral equation* for each agent \mathcal{A}_i , induced by his/her environmental psycho-physical field Φ_i , reads

$$\dot{\pi}_i = e_i \Phi_i \dot{\mathbf{q}}_i, \quad \text{in components,} \quad \dot{\pi}_\mu^i = e_i \Phi_{\mu\eta}^i \dot{q}_i^\eta.$$

This equation states that the force of an individual agent's motion, or his/her behavior, equals the product of his/her psycho-physical charge, environmental psycho-physical field, and velocity (speed) of behavior.

4.9.8.3 Collective Crowd Geometrodynamics

Now we define the total crowd's geometrodynamics as a union $\mathcal{C} = \cup_i \mathcal{A}_i$ of all individual agents' geometrodynamics, to model their emotion-induced behavior. For this we use the total *crowd's energy-momentum tensor* (CEM) and its divergence equation of motion.

As a union of individual Riemannian manifolds, the total crowd's manifold $\mathcal{M} = \cup_i M_i$ is also Riemannian, with the metrics $\mathbf{g} = \sum_i \mathbf{g}^i$ equal to the sum of individual metrics $\mathbf{g}^i = g_{\mu\eta}^i \omega_i^\mu \otimes \omega_i^\eta$. Now, the crowd's CEM tensor $\mathbf{C} = C_{\mu\eta} \omega_{\mathcal{M}}^\mu \otimes \omega_{\mathcal{M}}^\eta = \frac{1}{2} C_{\mu\eta} \omega_{\mathcal{M}}^\mu \wedge \omega_{\mathcal{M}}^\eta$ (where $\omega_{\mathcal{M}}^\mu$ denote the basis one-forms on \mathcal{M}) has two parts, $\mathbf{C} = \mathbf{C}^{(E)} + \mathbf{C}^{(B)}$, in components, $C_{\mu\eta} = C_{\mu\eta}^{(E)} + C_{\mu\eta}^{(B)}$, corresponding to the total crowd's emotion and behavior, which we define respectively as follows:

- (1) The CEM's *emotional part* $\mathbf{C}^{(E)}$ is in components defined as a sum of individual agents' motivational-tension fields,

$$C_{\mu\eta}^{(E)} = \sum_{i=1}^M (\Phi_{\mu\nu}^i \Phi_{i\eta}^\nu - g_{\mu\eta}^i \Phi_{\nu\lambda}^i \Phi_i^{\nu\lambda}),$$

so the equation of the crowd's emotion is defined in the form of the divergence equation

$$C_{\mu\eta}^{(E);\eta} = \sum_{i=1}^M \Phi_{\mu\eta}^i \Phi_{i\nu}^{\eta;\nu} = - \sum_{i=1}^M \Phi_{\mu\eta}^i J_i^\eta.$$

This equation says that CEM's *emotional part* $\mathbf{C}^{(E)}$ is a collective motivational field with a *sink*.

- (2) The CEM's *behavioral part* $\mathbf{C}^{(B)}$ as a sum of individual agents' motivational-currents times momenta:

$$C_{\mu\eta}^{(B)} = \sum_{i=1}^M g_{\mu\eta}^i \int_{t_i} \pi_{\eta}^i(t_i) \dot{q}_i^{\eta}(t_i) \delta^n[q_i^{\mu}(t_i)] dt_i,$$

so the equation of the crowd's behavior is defined in the form of the divergence equation

$$\begin{aligned} C_{\mu\eta}^{(B);\eta} &= \sum_{i=1}^M g_{\mu\eta}^i \int_{t_i} \pi_{\eta}^i(t_i) \dot{q}_i^{\eta}(t_i) \partial_{q_i^{\eta}} \delta^n[q_i^{\mu}(t_i)] dt_i \\ &= \sum_{i=1}^M e_i \int_{t_i} \Phi_{\mu\eta}^i \dot{q}_i^{\eta} \delta^n[q_i^{\mu}(t_i)] dt_i = \sum_{i=1}^M \Phi_{\mu\eta}^i J_i^{\eta}. \end{aligned}$$

This equation says that CEM's *behavioral part* $\mathbf{C}^{(B)}$ is a collective behavioral field with a *source*.

Therefore, the divergence equation for the total crowd's CEM tensor, represents the crowd's *motivation-behavior conservation law*

$$\operatorname{div} \mathbf{C} = \operatorname{div} \left(\mathbf{C}^{(E)} + \mathbf{C}^{(B)} \right) = 0, \quad \text{in components,} \quad C_{\mu\eta}^{(E);\eta} + C_{\mu\eta}^{(B);\eta} = 0.$$

This gives our basic representation of an isolated crowd as a conservative spatio-temporal dynamical system. Naturally, additional crowd's energy-momentum sources and sinks can violate this basic motivational-behavior conservation law.

4.10 Multivector-Fields and Tangent-Valued Forms

Recall that a vector-field on a manifold M is defined as a global section of the tangent bundle $TM \rightarrow M$. The set $\mathcal{V}^1(M)$ of vector-fields on M is a real Lie algebra with respect to the *Lie bracket* [Sardanashvily (1993); Sardanashvily (1995); Giachetta *et. al.* (1997); Mangiarotti and Sardanashvily (2000a); Sardanashvily (2002a)]

$$[v, u] = (v^{\alpha} \partial_{\alpha} u^{\mu} - u^{\alpha} \partial_{\alpha} v^{\mu}) \partial_{\mu}, \quad v = v^{\alpha} \partial_{\alpha}, \quad u = u^{\alpha} \partial_{\alpha}. \quad (4.123)$$

Every vector-field on a manifold M can be seen as an infinitesimal generator of a local 1-parameter Lie group of diffeomorphisms of M as follows [Kobayashi and Nomizu (1963/9)]. Given an open subset $U \subset M$ and an

interval $(-\epsilon, \epsilon) \in \mathbb{R}$, by a local 1-parameter group of diffeomorphisms of M defined on $(-\epsilon, \epsilon) \times U$ is denoted a map

$$G \rightarrow M, \quad (t, x) \mapsto G_t(x)$$

such that:

- (1) for each $t \in (-\epsilon, \epsilon)$, the map G_t is a diffeomorphism of U onto the open subset $G_t(U) \subset M$; and
- (2) $G_{t+t'}(x) = (G_t \circ G_{t'})(x)$ if $t, t', t + t' \in (-\epsilon, \epsilon)$ and $G_{t'}(x), x \in U$.

Any local 1-parameter group of diffeomorphisms G on $U \subset M$ defines a local vector-field u on U by setting $u(x)$ to be the tangent vector to the curve $x(t) = G_t(x)$ at $t = 0$. Conversely, if u is a vector-field on a manifold M , there exists a unique local 1-parameter group G_u of diffeomorphisms on a neighborhood of every point $x \in M$ which defines u . We call G_u a *flow* of the vector-field u . A vector-field u on a manifold M is called *complete* if its flow is a 1-parameter group of diffeomorphisms of M . In particular, every vector-field on a compact manifold is complete [Kobayashi and Nomizu (1963/9)].

A vector-field u on a fibre bundle $Y \rightarrow X$ is an infinitesimal generator of a local 1-parameter group G_u of isomorphisms of $Y \rightarrow X$ iff it is a projectable vector-field on Y . A vector-field u on a fibre bundle $Y \rightarrow X$ is called projectable if it projects onto a vector-field on X , i.e., there exists a vector-field τ on X such that the following diagram commutes:

$$\begin{array}{ccc} Y & \xrightarrow{u} & TY \\ \pi \downarrow & & \downarrow T\pi \\ X & \xrightarrow{\tau} & TX \end{array}$$

A projectable vector-field has the coordinate expression

$$u = u^\alpha(x^\mu) \partial_\alpha + u^i(x^\mu, y^j) \partial_i,$$

where u^α are local functions on X . A projectable vector-field is said to be vertical if it projects onto the zero vector-field $\tau = 0$ on X , i.e., $u = u^i \partial_i$ takes its values in the vertical tangent bundle VY .

For example, in field theory, projectable vector-fields on fibre bundles play a role of infinitesimal generators of local 1-parameter groups of gauge transformations.

In general, a vector-field $\tau = \tau^\alpha \partial_\alpha$ on a base X of a fibre bundle $Y \rightarrow X$ induces a vector-field on Y by means of a *connection* on this fibre bundle. Nevertheless, every natural fibre bundle $Y \rightarrow X$ admits the *canonical lift* $\tilde{\tau}$ onto Y of any vector-field τ on X . For example, if Y is the tensor bundle (4.11), the above canonical lift reads

$$\tilde{\tau} = \tau^\mu \partial_\mu + [\partial_\nu \tau^{\alpha_1} \dot{x}_{\beta_1 \dots \beta_k}^{\nu \alpha_2 \dots \alpha_m} + \dots - \partial_{\beta_1} \tau^\nu \dot{x}_{\nu \beta_2 \dots \beta_k}^{\alpha_1 \dots \alpha_m} - \dots] \frac{\partial}{\partial \dot{x}_{\beta_1 \dots \beta_k}^{\alpha_1 \dots \alpha_m}}. \quad (4.124)$$

In particular, we have the canonical lift onto the tangent bundle TX ,

$$\tilde{\tau} = \tau^\mu \partial_\mu + \partial_\nu \tau^\alpha \dot{x}^\nu \frac{\partial}{\partial \dot{x}^\alpha} \quad (4.125)$$

and another one onto the cotangent bundle T^*X ,

$$\tilde{\tau} = \tau^\mu \partial_\mu - \partial_\beta \tau^\nu \dot{x}_\nu \frac{\partial}{\partial \dot{x}_\beta}. \quad (4.126)$$

A *multivector-field* ϑ of degree r (or simply a r -vector-field) on a manifold M , by definition, is a global section of the bundle $\wedge^r TM \rightarrow M$. It is given by the coordinate expression

$$\vartheta = \vartheta^{\alpha_1 \dots \alpha_r} \partial_{\alpha_1} \wedge \dots \wedge \partial_{\alpha_r}, \quad |\vartheta| = r,$$

where summation is over all ordered collections $(\lambda_1, \dots, \lambda_r)$.

Similarly, an *exterior r -form* on a manifold M with local coordinates x^α , by definition, is a global section of the skew-symmetric tensor bundle (exterior product) $\wedge^r T^*M \rightarrow M$,

$$\phi = \frac{1}{r!} \phi_{\alpha_1 \dots \alpha_r} dx^{\alpha_1} \wedge \dots \wedge dx^{\alpha_r}, \quad |\phi| = r.$$

The 1-forms are also called the *Pfaffian forms*.

The vector space $\mathcal{V}^r(M)$ of r -vector-fields on a manifold M admits the *Schouten–Nijenhuis bracket* (or, SN bracket)

$$[\cdot, \cdot]_{SN} : \mathcal{V}^r(M) \times \mathcal{V}^s(M) \rightarrow \mathcal{V}^{r+s-1}(M)$$

which generalizes the Lie bracket of vector-fields (4.123). The SN-bracket has the coordinate expression:

$$\begin{aligned} \vartheta &= \vartheta^{\alpha_1 \dots \alpha_r} \partial_{\alpha_1} \wedge \dots \wedge \partial_{\alpha_r}, & v &= v^{\alpha_1 \dots \alpha_s} \partial_{\alpha_1} \wedge \dots \wedge \partial_{\alpha_s}, \\ [\vartheta, v]_{SN} &= \vartheta \star v + (-1)^{|\vartheta||v|} v \star \vartheta, & \text{where} \\ \vartheta \star v &= \vartheta^{\mu \alpha_1 \dots \alpha_{r-1}} \partial_\mu v^{\alpha_1 \dots \alpha_s} \partial_{\alpha_1} \wedge \dots \wedge \partial_{\alpha_{r-1}} \wedge \partial_{\alpha_1} \wedge \dots \wedge \partial_{\alpha_s}. \end{aligned}$$

The following relations hold for the SN-bracket:

$$\begin{aligned} [\vartheta, v]_{SN} &= (-1)^{|\vartheta||v|} [v, \vartheta]_{SN}, \\ [\nu, \vartheta \wedge v]_{SN} &= [\nu, \vartheta]_{SN} \wedge v + (-1)^{|\nu||\vartheta|+|\vartheta|} \vartheta \wedge [\nu, v]_{SN}, \\ (-1)^{|\nu||\vartheta|+|\nu|} [\nu, \vartheta \wedge v]_{SN} &+ (-1)^{|\vartheta||\nu|+|\vartheta|} [\vartheta, v \wedge \nu]_{SN} \\ &+ (-1)^{|\nu||\vartheta|+|\nu|} [v, \nu \wedge \vartheta]_{SN} = 0. \end{aligned}$$

In particular, let $w = w^{\mu\nu} \partial_\mu \wedge \partial_\nu$ be a *bivector-field*. We have

$$[w, w]_{SN} = w^{\mu\alpha_1} \partial_\mu w^{\alpha_2\alpha_3} \partial_{\alpha_1} \wedge \partial_{\alpha_2} \wedge \partial_{\alpha_3}. \quad (4.127)$$

Every bivector-field w on a manifold M induces the ‘sharp’ bundle map $w^\sharp : T^*M \rightarrow TM$ defined by

$$w^\sharp(p)]q := w(x)(p, q), \quad w^\sharp(p) = w^{\mu\nu}(x) p_\mu \partial_\nu, \quad (p, q \in T_x^*M). \quad (4.128)$$

A bivector-field w whose bracket (4.127) vanishes is called the *Poisson bivector-field*.

Let $\wedge^r(M)$ denote the vector space of exterior r -forms on a manifold M . By definition, $\wedge^0(M) = C^\infty(M)$ is the ring of smooth real functions on M . All exterior forms on M constitute the \mathbb{N} -graded exterior algebra $\wedge^*(M)$ of global sections of the exterior bundle $\wedge T^*M$ with respect to the exterior product \wedge . This algebra admits the *exterior differential*

$$\begin{aligned} d : \wedge^r(M) &\rightarrow \wedge^{r+1}(M), \\ d\phi &= dx^\mu \wedge \partial_\mu \phi = \frac{1}{r!} \partial_\mu \phi_{\alpha_1 \dots \alpha_r} dx^\mu \wedge dx^{\alpha_1} \wedge \dots \wedge dx^{\alpha_r}, \end{aligned}$$

which is *nilpotent*, i.e., $d \circ d = 0$, and obeys the relation

$$d(\phi \wedge \sigma) = d(\phi) \wedge \sigma + (-1)^{|\phi|} \phi \wedge d(\sigma).$$

The *interior product* (or, *contraction*) of a vector-field $u = u^\mu \partial_\mu$ and an exterior r -form ϕ on a manifold M is given by the coordinate expression

$$\begin{aligned} u] \phi &= \sum_{k=1}^r \frac{(-1)^{k-1}}{r!} u^{\alpha_k} \phi_{\alpha_1 \dots \alpha_k \dots \alpha_r} dx^{\alpha_1} \wedge \dots \wedge \widehat{dx^{\alpha_k}} \wedge \dots \wedge dx^{\alpha_r} \\ &= \frac{1}{(r-1)!} u^\mu \phi_{\mu \alpha_2 \dots \alpha_r} dx^{\alpha_2} \wedge \dots \wedge dx^{\alpha_r}, \end{aligned} \quad (4.129)$$

where the caret $\hat{\cdot}$ denotes omission. The following relations hold:

$$\phi(u_1, \dots, u_r) = u_r \rfloor \cdots u_1 \rfloor \phi, \quad (4.130)$$

$$u \rfloor (\phi \wedge \sigma) = u \rfloor \phi \wedge \sigma + (-1)^{|\phi|} \phi \wedge u \rfloor \sigma, \quad (4.131)$$

$$[u, u'] \rfloor \phi = u \rfloor d(u' \rfloor \phi) - u' \rfloor d(u \rfloor \phi) - u' \rfloor u \rfloor d\phi, \quad (\phi \in \wedge^1(M)). \quad (4.132)$$

Recall from section 3.7 above, that the *Lie derivative* $\mathfrak{L}_u \sigma$ of an exterior form σ along a vector-field u is defined by the *Cartan relation*

$$\mathfrak{L}_u \sigma = u \rfloor d\sigma + d(u \rfloor \sigma).$$

It satisfies the relation

$$\mathfrak{L}_u(\phi \wedge \sigma) = \mathfrak{L}_u \phi \wedge \sigma + \phi \wedge \mathfrak{L}_u \sigma.$$

In particular, if f is a function, then

$$\mathfrak{L}_u f = u(f) = u \rfloor df.$$

It is important for dynamical applications that an exterior form ϕ is invariant under a local 1-parameter group of diffeomorphisms G_t of M (i.e., $G_t^* \phi = \phi$) iff its Lie derivative $\mathfrak{L}_u \phi$ along the vector-field u , generating G_t , vanishes.

Let Ω be a two-form on M . It defines the ‘flat’ bundle map Ω^\flat , as

$$\Omega^\flat : TM \rightarrow T^*M, \quad \Omega^\flat(v) = -v \rfloor \Omega(x), \quad (v \in T_x M). \quad (4.133)$$

In coordinates, if $\Omega = \Omega_{\mu\nu} dx^\mu \wedge dx^\nu$ and $v = v^\mu \partial_\mu$, then

$$\Omega^\flat(v) = -\Omega_{\mu\nu} v^\mu dx^\nu.$$

One says that Ω is of *constant rank* k if the corresponding map (4.133) is of constant rank k (i.e., k is the greatest integer n such that Ω^n is not the zero form). The rank of a *nondegenerate* two-form is equal to $\dim M$. A nondegenerate closed two-form is called the *symplectic form*.

Given a manifold map $f : M \rightarrow M'$, any exterior k -form ϕ on M' induces the pull-back exterior form $f^* \phi$ on M by the condition

$$f^* \phi(v^1, \dots, v^k)(x) = \phi(Tf(v^1), \dots, Tf(v^k))(f(x))$$

for an arbitrary collection of tangent vectors $v^1, \dots, v^k \in T_x M$. The following relations hold:

$$f^*(\phi \wedge \sigma) = f^* \phi \wedge f^* \sigma, \quad df^* \phi = f^*(d\phi).$$

In particular, given a fibre bundle $\pi : Y \rightarrow X$, the pull-back onto Y of exterior forms on X by π gives the monomorphism of exterior algebras

$$\pi^* : \wedge^*(X) \rightarrow \wedge^*(Y).$$

Elements of its image $\pi^* \wedge^*(X)$ are called basic forms. Exterior forms on Y such that $u \rfloor \phi = 0$ for an arbitrary vertical vector-field u on Y are said to be *horizontal forms*. They are generated by horizontal 1-forms $\{dx^\alpha\}$. For example, basic forms are horizontal forms with coefficients in $C^\infty(X) \subset C^\infty(Y)$. A horizontal form of degree $n = \dim X$ is called a *density*. For example, Lagrangians in field theory are densities.

Elements of the tensor product $\wedge^r(M) \otimes \mathcal{V}^1(M)$ are called the *tangent-valued r -forms* on M . They are sections

$$\phi = \frac{1}{r!} \phi_{\alpha_1 \dots \alpha_r}^\mu dx^{\alpha_1} \wedge \dots \wedge dx^{\alpha_r} \otimes \partial_\mu$$

of the tensor bundle

$$\wedge^r T^*M \otimes TM \rightarrow M.$$

Tangent-valued 1-forms are usually called the (1,1) *tensor-fields*.

In particular, there is the 1-1 correspondence between the tangent-valued 1-forms on M and the linear bundle maps over M ,

$$\phi : TM \rightarrow TM, \quad \phi : T_x M \ni v \mapsto v \rfloor \phi(x) \in T_x M. \quad (4.134)$$

In particular, the *canonical tangent-valued one-form* $\theta_M = dx^\alpha \otimes \partial_\alpha$ defines the identity map of TM .

Tangent-valued forms play a prominent role in jet formalism and theory of connections on fibre bundles. In particular, tangent-valued 0-forms are vector-fields on M . Also, there is 1-1 correspondence between the tangent-valued 1-forms ϕ on a manifold M and the linear bundle endomorphisms

$$\hat{\phi} : TM \rightarrow TM, \quad \hat{\phi} : T_x M \ni v \mapsto v \rfloor \phi(x) \in T_x M, \quad (4.135)$$

$$\hat{\phi}^* : T^*M \rightarrow T^*M, \quad \hat{\phi}^* : T_x^*M \ni v^* \mapsto \phi(x) \rfloor v^* \in T_x^*M, \quad (4.136)$$

over M . For example, the *canonical tangent-valued 1-form* on M ,

$$\theta_M = dx^\alpha \otimes \partial_\alpha, \quad (4.137)$$

corresponds to the identity maps (4.135) and (4.136).

We shall deal with the following particular types of vector-fields and differential forms on a bundle $Y \rightarrow X$ [Sardanashvily (1993); Sardanashvily (1995); Giachetta *et. al.* (1997); Mangiarotti and Sardanashvily (2000a)]:

- a *projectable vector-field* on Y ,

$$u = u^\mu(x)\partial_\mu + u^i(y)\partial_i,$$

which covers a vector-field $\tau_u = u^\mu(x)\partial_\mu$ on the base X such that the following diagram commutes:

$$\begin{array}{ccc} Y & \xrightarrow{u} & TY \\ \pi \downarrow & & \downarrow T\pi \\ X & \xrightarrow{\tau_u} & TX \end{array}$$

- a *vertical vector-field*, $u : Y \rightarrow TY$, given by $u = u^i(y)\partial_i$, is a projectable vector-field which covers $\tau_u = 0$;
- an *exterior horizontal form*, $\phi : Y \rightarrow \wedge^r T^*X$, given by

$$\phi = \frac{1}{r!} \phi_{\alpha_1 \dots \alpha_r}(y) dx^{\alpha_1} \wedge \dots \wedge dx^{\alpha_r};$$

- a *tangent-valued horizontal form*, $\phi : Y \rightarrow \wedge^r T^*X \otimes TY$, given by

$$\phi = \frac{1}{r!} dx^{\alpha_1} \wedge \dots \wedge dx^{\alpha_r} \otimes [\phi_{\alpha_1 \dots \alpha_r}^\mu(y)\partial_\mu + \phi_{\alpha_1 \dots \alpha_r}^i(y)\partial_i];$$

- a *vertical-valued horizontal form*, $\phi : Y \rightarrow \wedge^r T^*X \otimes VY$, given by

$$\phi = \frac{1}{r!} \phi_{\alpha_1 \dots \alpha_r}^i(y) dx^{\alpha_1} \wedge \dots \wedge dx^{\alpha_r} \otimes \partial_i.$$

- a vertical-valued *soldering form*, $\sigma : Y \rightarrow T^*X \otimes VY$, given by

$$\sigma = \sigma_\alpha^i(y) dx^\alpha \otimes \partial_i \quad (4.138)$$

and, in particular, the *canonical soldering form* on TX ,

$$\theta_X = dx^\alpha \otimes \partial_\alpha.$$

The *pull-back-valued forms* on a bundle $Y \rightarrow X$ are the following two maps:¹⁵

$$\begin{aligned} Y &\rightarrow \wedge^r T^*Y \otimes TX, & \phi &= \frac{1}{r!} \phi_{\alpha_1 \dots \alpha_r}^\mu(y) dx^{\alpha_1} \wedge \dots \wedge dx^{\alpha_r} \otimes \partial_\mu, \\ &\text{and} & & \\ Y &\rightarrow \wedge^r T^*Y \otimes V^*X, & \phi &= \frac{1}{r!} \phi_{\alpha_1 \dots \alpha_r i}(y) dx^{\alpha_1} \wedge \dots \wedge dx^{\alpha_r} \otimes \bar{d}y^i. \end{aligned} \quad (4.139)$$

The pull-back-valued forms (4.139) are exemplified by the canonical bundle monomorphism

$$\wedge^n T^*X \otimes V^*Y \hookrightarrow \wedge^{n+1} T^*Y, \quad \omega \otimes \bar{d}y^i \mapsto \omega \wedge dy^i.$$

All horizontal n -forms on a bundle $Y \rightarrow X$ are called *horizontal densities*.

For any vector-field τ on X , we can define its pull-back on Y ,

$$\pi^* \tau = \tau \circ \pi : Y \rightarrow TX.$$

This is not a vector-field on Y , for the tangent bundle TX of X fails to be a subbundle of the tangent bundle TY of Y . One needs a connection on $Y \rightarrow X$ in order to set the imbedding $TX \hookrightarrow TY$.

The space $\wedge^*(M) \otimes \mathcal{V}^1(M)$ of tangent-valued forms admits the *Frölicher–Nijenhuis bracket* (or, FN bracket)

$$\begin{aligned} [\cdot, \cdot]_{FN} : \wedge^r(M) \otimes \mathcal{V}^1(M) \times \wedge^s(M) \otimes \mathcal{V}^1(M) &\rightarrow \wedge^{r+s}(M) \otimes \mathcal{V}^1(M), \\ [\phi, \sigma]_{FN} &= \frac{1}{r!s!} (\phi_{\alpha_1 \dots \alpha_r}^\nu \partial_\nu \sigma_{\alpha_{r+1} \dots \alpha_{r+s}}^\mu - \sigma_{\alpha_{r+1} \dots \alpha_{r+s}}^\nu \partial_\nu \phi_{\alpha_1 \dots \alpha_r}^\mu - \\ &r \phi_{\alpha_1 \dots \alpha_{r-1} \nu}^\mu \partial_\alpha \sigma_{\alpha_{r+1} \dots \alpha_{r+s}}^\nu + s \sigma_{\nu \alpha_{r+2} \dots \alpha_{r+s}}^\mu \partial_{\alpha_{r+1}} \phi_{\alpha_1 \dots \alpha_r}^\nu) dx^{\alpha_1} \\ &\wedge \dots \wedge dx^{\alpha_{r+s}} \otimes \partial_\mu. \end{aligned} \quad (4.140)$$

¹⁵The forms (4.139) are not tangent-valued forms. The pull-backs

$$\phi = \frac{1}{r!} \phi_{\alpha_1 \dots \alpha_r}^\mu(x) dx^{\alpha_1} \wedge \dots \wedge dx^{\alpha_r} \otimes \partial_\mu$$

of tangent-valued forms on X onto Y by π exemplify the pull-back-valued forms (4.139). In particular, we shall refer to the pull-back $\pi^* \theta_X$ of the canonical form θ_X on the base X onto Y by π . This is a pull-back-valued horizontal one-form on Y which we denote by the same symbol

$$\theta_X : Y \rightarrow T^*X \otimes TX, \quad \theta_X = dx^\alpha \otimes \partial_\alpha.$$

The following relations hold for the FN-bracket:

$$\begin{aligned} [\phi, \psi]_{FN} &= (-1)^{|\phi||\psi|+1} [\psi, \phi]_{FN}, \\ [\phi, [\psi, \theta]_{FN}]_{FN} &= [[\phi, \psi]_{FN}, \theta]_{FN} + (-1)^{|\phi||\psi|} [\psi, [\phi, \theta]_{FN}]_{FN}. \end{aligned} \quad (4.141)$$

Given a tangent-valued form θ , the *Nijenhuis differential*, $d_\theta \sigma$, along θ on $\wedge^*(M) \otimes \mathcal{V}^1(M)$ is defined as

$$d_\theta \sigma = [\theta, \sigma]_{FN}. \quad (4.142)$$

By virtue of the relation (4.141), it has the property

$$d_\phi [\psi, \theta]_{FN} = [d_\phi \psi, \theta]_{FN} + (-1)^{|\phi||\psi|} [\psi, d_\phi \theta]_{FN}.$$

In particular, if $\theta = u$ is a vector-field, the Nijenhuis differential becomes the Lie derivative of tangent-valued forms

$$\begin{aligned} \mathcal{L}_u \sigma &= d_u \sigma = [u, \sigma]_{FN} = (u^\nu \partial_\nu \sigma_{\alpha_1 \dots \alpha_s}^\mu - \sigma_{\alpha_1 \dots \alpha_s}^\nu \partial_\nu u^\mu \\ &\quad + s \sigma_{\nu \alpha_2 \dots \alpha_s}^\mu \partial_{\alpha_1} u^\nu) dx^{\alpha_1} \wedge \dots \wedge dx^{\alpha_s} \otimes \partial_\mu, \quad (\sigma \in \wedge^s(M) \otimes \mathcal{V}(M)). \end{aligned} \quad (4.143)$$

4.11 Application: Geometrical Quantization

4.11.1 Quantization of Hamiltonian Mechanics

Recall that classical *Dirac quantization* states [Dirac (1982)]:

$$\{f, g\} = \frac{1}{i\hbar} [\hat{f}, \hat{g}],$$

which means that the quantum Poisson brackets (i.e., commutators) have the same values as the classical Poisson brackets. In other words, we can associate smooth functions defined on the symplectic phase-space manifold (M, ω) of the classical biodynamic system with operators on a Hilbert space \mathcal{H} in such a way that the Poisson brackets correspond. Therefore, *there is a functor* from the category **Symplec** to the category **Hilbert**. This functor is called *prequantization*.

Let us start with the simplest symplectic manifold $(M = T^*\mathbb{R}^n, \omega = dp_i \wedge dq^i)$ and state the *Dirac problem*: A prequantization of $(T^*\mathbb{R}^n, \omega = dp_i \wedge dq^i)$ is a map $\delta : f \mapsto \delta_f$, taking smooth functions $f \in C^\infty(T^*\mathbb{R}^n, \mathbb{R})$ to Hermitian operators δ_f on a Hilbert space \mathcal{H} , satisfying the *Dirac conditions*:

- (1) $\delta_{f+g} = \delta_f + \delta_g$, for each $f, g \in C^\infty(T^*\mathbb{R}^n, \mathbb{R})$;

- (2) $\delta_{\lambda f} = \lambda \delta_f$, for each $f \in C^\infty(T^*\mathbb{R}^n, \mathbb{R})$ and $\lambda \in \mathbb{R}$;
 (3) $\delta_{1_{\mathbb{R}^n}} = Id_{\mathcal{H}}$; and
 (4) $[\delta_f, \delta_g] = (\delta_f \circ \delta_g - \delta_g \circ \delta_f) = i\hbar \delta_{\{f, g\}_\omega}$, for each $f, g \in C^\infty(T^*\mathbb{R}^n, \mathbb{R})$;

The pair (\mathcal{H}, δ) , where

$$\begin{aligned} \mathcal{H} &= L^2(\mathbb{R}^n, \mathbb{C}); & \delta : f \in C^\infty(T^*\mathbb{R}^n, \mathbb{R}) &\mapsto \delta_f : \mathcal{H} \rightarrow \mathcal{H}; \\ \delta_f &= -i\hbar X_f - \theta(X_f) + f; & \theta &= p_i dq^i, \end{aligned}$$

gives a prequantization of $(T^*\mathbb{R}^n, dp_i \wedge dq^i)$, or equivalently, the answer to the Dirac problem is affirmative [Putz (1993)].

Now, let $(M = T^*Q, \omega)$ be the cotangent bundle of an arbitrary manifold Q with its canonical symplectic structure $\omega = d\theta$. The prequantization of M is given by the pair $(L^2(M, \mathbb{C}), \delta^\theta)$, where for each $f \in C^\infty(M, \mathbb{R})$, the operator $\delta_f^\theta : L^2(M, \mathbb{C}) \rightarrow L^2(M, \mathbb{C})$ is given by

$$\delta_f^\theta = -i\hbar X_f - \theta(X_f) + f.$$

Here, *symplectic potential* θ is not uniquely determined by the condition $\omega = d\theta$; for instance $\theta' = \theta + du$ has the same property for any real function u on M . On the other hand, in the general case of an arbitrary symplectic manifold (M, ω) (not necessarily the cotangent bundle) we can find only locally a 1-form θ such that $\omega = d\theta$.

In general, a symplectic manifold $(M, \omega = d\theta)$ is quantizable (i.e., we can define the Hilbert representation space \mathcal{H} and the prequantum operator δ_f in a globally consistent way) if ω defines an *integral cohomology class*. Now, by the *construction Theorem of a fiber bundle*, we can see that this condition on ω is also sufficient to guarantee the existence of a *complex line bundle* $L^\omega = (L, \pi, M)$ over M , which has $\exp(i u_{ji}/\hbar)$ as *gauge transformations* associated to an open cover $\mathcal{U} = \{U_i | i \in I\}$ of M such that θ_i is a *symplectic potential* defined on U_i (i.e., $d\theta_i = \omega$ and $\theta_i = \theta_j + du_{ji}$ on $U_i \cap U_j$).

In particular, for exact symplectic structures ω (as in the case of cotangent bundles with their canonical symplectic structures) an integral cohomology condition is automatically satisfied, since then we have only one set $U_i = M$ and do not need any gauge transformations.

Now, for each vector-field $X \in M$ there exists an operator ∇_X^ω on the space of sections $\Gamma(L^\omega)$ of L^ω ,

$$\nabla_X^\omega : \Gamma(L^\omega) \rightarrow \Gamma(L^\omega), \quad \text{given by} \quad \nabla_X^\omega f = X(f) - \frac{i}{\hbar} \theta(X)f,$$

and it is easy to see that ∇^ω is a *connection* on L^ω whose curvature is $\omega/i\hbar$. In terms of this connection, the definition of δ_f becomes

$$\delta_f = -i\hbar \nabla_{X_f}^\omega + f.$$

The complex line bundle $L^\omega = (L, \pi, M)$ together with its compatible connection and Hermitian structure is usually called the *prequantum bundle* of the symplectic manifold (M, ω) .

If (M, ω) is a quantizable manifold then the pair (\mathcal{H}, δ) defines its prequantization.

Examples

Each exact symplectic manifold $(M, \omega = d\theta)$ is quantizable, for the cohomology class defined by ω is zero. In particular, the cotangent bundle, with its canonical symplectic structure is always quantizable.

Let $(M, \omega = d\theta)$ be an exact symplectic manifold. Then it is quantizable with the prequantum bundle given by [Puta (1993)]:

$$\begin{aligned} L^\omega &= (M \times \mathbb{C}, pr_1, M); \\ \Gamma(L^\omega) &\simeq C^\infty(M, \mathbb{C}); \quad \nabla_X^\omega f = X(f) - \frac{i}{\hbar} \theta(X)f; \\ ((x, z_1), (x, z_2))_x &= \bar{z}_1 z_2; \quad \delta_f = -i\hbar [X_f - \frac{i}{\hbar} \theta(X_f)] + f. \end{aligned}$$

Let $(M, \omega) = (T^*\mathbb{R}, dp \wedge dq)$. It is quantizable with [Puta (1993)]:

$$\begin{aligned} L^\omega &= (\mathbb{R}^2 \times \mathbb{C}, pr_1, \mathbb{R}^2); \quad \Gamma(L^\omega) = C^\infty(\mathbb{R}^2, \mathbb{C}); \\ \nabla_X^\omega f &= X(f) - \frac{i}{\hbar} pdq(X)f; \quad ((x, z_1), (x, z_2))_x = \bar{z}_1 z_2; \\ \delta_f &= -i\hbar \left[\frac{\partial f}{\partial p} \frac{\partial}{\partial q} - \frac{\partial f}{\partial q} \frac{\partial}{\partial p} \right] - p \frac{\partial f}{\partial p} + f. \end{aligned}$$

Therefore,

$$\delta_q = i\hbar \frac{\partial}{\partial p} + q, \quad \delta_p = -i\hbar \frac{\partial}{\partial q},$$

which differs from the classical result of the *Schrödinger quantization*:

$$\delta_q = q, \quad \delta_p = -i\hbar \frac{\partial}{\partial q}.$$

Let \mathcal{H} be a complex Hilbert space and $U_t : \mathcal{H} \rightarrow \mathcal{H}$ a continuous one-parameter unitary group, i.e., a homomorphism $t \mapsto U_t$ from \mathbb{R} to the group

of unitary operators on \mathcal{H} such that for each $x \in \mathcal{H}$ the map $t \mapsto U_t(x)$ is continuous. Then we have the self-adjoint generator A of U_t , defined by

$$Ax = \frac{1}{i} \frac{d}{dt} U_t(x) = \frac{1}{i} \lim_{h \rightarrow 0} \frac{U_h(x) - x}{h}.$$

Let $(\mathbb{R}^2, \omega = dp \wedge dq, H = \frac{1}{2}(p^2 + q^2))$ be the Hamiltonian structure of the 1D harmonic oscillator.

If we take $\theta = \frac{1}{2}(pdq - qdp)$ as the symplectic potential of ω , then the spectrum of the prequantum operator $\delta_H = i\hbar \left(q \frac{\partial}{\partial p} - p \frac{\partial}{\partial q} \right)$ is [Puta (1993)]

$\text{Spec}(\delta_H) = \{\dots, -2\hbar, -\hbar, 0, \hbar, 2\hbar, \dots\}$, where each eigenvalue occurs with infinite multiplicity.

Let \mathfrak{g} be the vector space spanned by the prequantum operators $\delta_q, \delta_p, \delta_H$, given by

$$\delta_q = i\hbar \frac{\partial}{\partial p} + q, \quad \delta_p = -i\hbar \frac{\partial}{\partial q}, \quad \delta_H = i\hbar \left(q \frac{\partial}{\partial p} - p \frac{\partial}{\partial q} \right),$$

and Id . Then we have [Puta (1993)]:

(1) \mathfrak{g} is a Lie algebra called the *oscillator Lie algebra*, given by:

$$\begin{aligned} [\delta_p, \delta_q] &= i\hbar \delta_{\{p, q\}_\omega} = i\hbar Id, \\ [\delta_H, \delta_q] &= i\hbar \delta_{\{H, q\}_\omega} = -i\hbar \delta_p, \\ [\delta_H, \delta_p] &= i\hbar \delta_{\{H, p\}_\omega} = i\hbar \delta_q, \end{aligned}$$

(2) $[\mathfrak{g}, \mathfrak{g}]$ is spanned by $\delta_q, \delta_p, \delta_H$ and Id , or equivalently, it is a *Heisenberg Lie algebra*.

(3) The oscillator Lie algebra \mathfrak{g} is *solvable*.

4.11.2 Quantization of Relativistic Hamiltonian Mechanics

Given a symplectic manifold (Z, Ω) and a Hamiltonian H on Z , a Dirac constraint system on a closed imbedded submanifold $i_N : N \rightarrow Z$ of Z is defined as a Hamiltonian system on N admitting the pull-back presymplectic form $\Omega_N = i_N^* \Omega$ and the pull-back Hamiltonian $i_N^* H$ [Gotay *et. al.* (1978); Mangiarotti and Sardanashvily (1998); Muñoz and Román (1992)]. Its solution is a vector-field γ on N which fulfils the equation

$$\gamma \lrcorner \Omega_N + i_N^* dH = 0.$$

Let N be coisotropic. Then a solution exists if the Poisson bracket $\{H, f\}$ vanishes on N whenever f is a function vanishing on N . It is the Hamiltonian vector-field of H on Z restricted to N [Sardanashvily (2003)].

Recall that a configuration space of non-relativistic time-dependent mechanics (henceforth NRM) of m degrees of freedom is an $(m+1)$ D smooth fibre bundle $Q \rightarrow \mathbb{R}$ over the time axis \mathbb{R} [Mangiarotti and Sardanashvily (1998); Sardanashvily (1998)]. It is coordinated by $(q^\alpha) = (q^0, q^i)$, where $q^0 = t$ is the standard Cartesian coordinate on \mathbb{R} . Let T^*Q be the cotangent bundle of Q equipped with the induced coordinates $(q^\alpha, p_\alpha = \dot{q}_\alpha)$ with respect to the holonomic coframes $\{dq^\alpha\}$. The cotangent bundle T^*Q plays the role of a homogeneous momentum phase-space of NRM, admitting the canonical symplectic form

$$\Omega = dp_\alpha \wedge dq^\alpha. \quad (4.144)$$

Its momentum phase-space is the vertical cotangent bundle V^*Q of the configuration bundle $Q \rightarrow \mathbb{R}$, coordinated by (q^α, q^i) . A Hamiltonian \mathcal{H} of NRM is defined as a section $p_0 = -\mathcal{H}$ of the fibre bundle $T^*Q \rightarrow V^*Q$. Then the momentum phase-space of NRM can be identified with the image N of \mathcal{H} in T^*Q which is the one-codimensional (consequently, coisotropic) imbedded submanifold given by the constraint

$$\mathcal{H}_T = p_0 + \mathcal{H}(q^\alpha, p_k) = 0.$$

Furthermore, a solution of a non-relativistic Hamiltonian system with a Hamiltonian \mathcal{H} is the restriction γ to $N \cong V^*Q$ of the Hamiltonian vector-field of \mathcal{H}_T on T^*Q . It obeys the equation $\gamma| \Omega_N = 0$ [Mangiarotti and Sardanashvily (1998); Sardanashvily (1998)]. Moreover, one can show that geometrical quantization of V^*Q is equivalent to geometrical quantization of the cotangent bundle T^*Q where the quantum constraint $\hat{\mathcal{H}}_T \psi = 0$ on sections ψ of the quantum bundle serves as the Schrödinger equation [Sardanashvily (2003)].

A configuration space of *relativistic mechanics* (henceforth RM) is an oriented pseudo-Riemannian manifold (Q, g) , coordinated by (t, q^i) . Its momentum phase-space is the cotangent bundle T^*Q provided with the symplectic form Ω (4.144). Note that one also considers another symplectic form $\Omega + F$ where F is the strength of an electromagnetic field [Sniatycki (1980)]. A relativistic Hamiltonian is defined as a smooth real function H on T^*Q [Mangiarotti and Sardanashvily (1998); Sardanashvily (1998)]. Then a relativistic Hamiltonian system is described as a Dirac constraint

system on the subspace N of T^*Q given by the equation

$$H_T = g_{\mu\nu} \partial^\mu H \partial^\nu H - 1 = 0. \quad (4.145)$$

To perform geometrical quantization of NRM, we give geometrical quantization of the cotangent bundle T^*Q and characterize a quantum relativistic Hamiltonian system by the quantum constraint

$$\hat{H}_T \psi = 0. \quad (4.146)$$

We choose the vertical polarization on T^*Q spanned by the tangent vectors ∂^α . The corresponding *quantum algebra* $\mathcal{A} \subset C^\infty(T^*Q)$ consists of affine functions of momenta

$$f = a^\alpha(q^\mu) p_\alpha + b(q^\mu) \quad (4.147)$$

on T^*Q . They are represented by the *Schrödinger operators*

$$\hat{f} = -i a^\alpha \partial_\alpha - \frac{i}{2} \partial_\alpha a^\alpha - \frac{i}{4} a^\alpha \partial_\alpha \ln(-g) + b, \quad (g = \det(g_{\alpha\beta})) \quad (4.148)$$

in the space $\mathbb{C}^\infty(Q)$ of smooth complex functions on Q .

Note that the function H_T (4.145) need not belong to the quantum algebra \mathcal{A} . Nevertheless, one can show that, if H_T is a polynomial of momenta of degree k , it can be represented as a finite composition

$$H_T = \sum_i f_{1i} \cdots f_{ki} \quad (4.149)$$

of products of affine functions (4.147), i.e., as an element of the enveloping algebra $\overline{\mathcal{A}}$ of the Lie algebra \mathcal{A} [Giachetta *et. al.* (2002b)]. Then it is quantized

$$H_T \mapsto \hat{H}_T = \sum_i \hat{f}_{1i} \cdots \hat{f}_{ki} \quad (4.150)$$

as an element of $\overline{\mathcal{A}}$. However, the representation (4.149) and, consequently, the quantization (4.150) fail to be unique.

The space of relativistic velocities of RM on Q is the tangent bundle TQ of Q equipped with the induced coordinates (t, q^i, \dot{q}^α) with respect to the holonomic frames $\{\partial_\alpha\}$. Relativistic motion is located in the subbundle W_g of hyperboloids [Mangiarotti and Sardanashevily (1998); Mangiarotti and Sardanashevily (2000b)]

$$g_{\mu\nu}(q) \dot{q}^\mu \dot{q}^\nu - 1 = 0 \quad (4.151)$$

of TQ . It is described by a second-order dynamical equation

$$\ddot{q}^\alpha = \Xi^\alpha(q^\mu, \dot{q}^\mu) \quad (4.152)$$

on Q which preserves the subbundle (4.151), i.e.,

$$(\dot{q}^\alpha \partial_\alpha + \Xi^\alpha \dot{\partial}_\alpha)(g_{\mu\nu} \dot{q}^\mu \dot{q}^\nu - 1) = 0, \quad (\dot{\partial}_\alpha = \partial/\partial \dot{q}^\alpha).$$

This condition holds if the r.h.s. of the equation (4.152) takes the form

$$\Xi^\alpha = \Gamma_{\mu\nu}^\alpha \dot{q}^\mu \dot{q}^\nu + F^\alpha,$$

where $\Gamma_{\mu\nu}^\alpha$ are Christoffel symbols of a metric g , while F^α obey the relation $g_{\mu\nu} F^\mu \dot{q}^\nu = 0$. In particular, if the dynamical equation (4.152) is a geodesic equation,

$$\ddot{q}^\alpha = K_\mu^\alpha \dot{q}^\mu$$

with respect to a (non-linear) connection on the tangent bundle $TQ \rightarrow Q$,

$$K = dq^\alpha \otimes (\partial_\alpha + K_\mu^\alpha \dot{\partial}_\mu),$$

this connections splits into the sum

$$K_\mu^\alpha = \Gamma_{\mu\nu}^\alpha \dot{q}^\nu + F_\mu^\alpha \quad (4.153)$$

of the Levi-Civita connection of g and a soldering form

$$F = g^{\lambda\nu} F_{\mu\nu} dq^\mu \otimes \dot{\partial}_\lambda, \quad F_{\mu\nu} = -F_{\nu\mu}.$$

As was mentioned above, the momentum phase-space of RM on Q is the cotangent bundle T^*Q provided with the symplectic form Ω (4.144). Let H be a smooth real function on T^*Q such that the map

$$\tilde{H} : T^*Q \rightarrow TQ, \quad \dot{q}^\mu = \partial^\mu H \quad (4.154)$$

is a bundle isomorphism. Then the inverse image $N = \tilde{H}^{-1}(W_g)$ of the subbundle of hyperboloids W_g (4.151) is a one-codimensional (consequently, coisotropic) closed imbedded subbundle of T^*Q given by the constraint $H_T = 0$ (4.145). We say that H is a relativistic Hamiltonian if the Poisson bracket $\{H, H_T\}$ vanishes on N . This means that the Hamiltonian vector-field

$$\gamma = \partial^\alpha H \partial_\alpha - \partial_\alpha H \partial^\alpha \quad (4.155)$$

of H preserves the constraint N and, restricted to N , it obeys the Hamiltonian equation

$$\gamma| \Omega_N + i_N^* dH = 0 \quad (4.156)$$

of a Dirac constraint system on N with a Hamiltonian H .

The map (4.154) sends the vector-field γ (4.155) onto the vector-field

$$\gamma_T = \dot{q}^\alpha \partial_\alpha + (\partial^\mu H \partial^\alpha \partial_\mu H - \partial_\mu H \partial^\alpha \partial^\mu H) \dot{\partial}_\alpha$$

on TQ . This vector-field defines the second-order dynamical equation

$$\ddot{q}^\alpha = \partial^\mu H \partial^\alpha \partial_\mu H - \partial_\mu H \partial^\alpha \partial^\mu H \quad (4.157)$$

on Q which preserves the subbundle of hyperboloids (4.151).

The following is a basic example of relativistic Hamiltonian systems. Put

$$H = \frac{1}{2m} g^{\mu\nu} (p_\mu - b_\mu)(p_\nu - b_\nu),$$

where m is a constant and $b_\mu dq^\mu$ is a covector-field on Q . Then $H_T = 2m^{-1}H - 1$ and $\{H, H_T\} = 0$. The constraint $H_T = 0$ defines a closed imbedded one-codimensional subbundle N of T^*Q . The Hamiltonian equation (4.156) takes the form $\gamma| \Omega_N = 0$. Its solution (4.155) reads

$$\begin{aligned} \dot{q}^\alpha &= \frac{1}{m} g^{\alpha\nu} (p_\nu - b_\nu), \\ \dot{p}_\alpha &= -\frac{1}{2m} \partial_\alpha g^{\mu\nu} (p_\mu - b_\mu)(p_\nu - b_\nu) + \frac{1}{m} g^{\mu\nu} (p_\mu - b_\mu) \partial_\alpha b_\nu. \end{aligned}$$

The corresponding second-order dynamical equation (4.157) on Q is

$$\begin{aligned} \ddot{q}^\alpha &= \Gamma_{\mu\nu}^\alpha \dot{q}^\mu \dot{q}^\nu - \frac{1}{m} g^{\lambda\nu} F_{\mu\nu} \dot{q}^\mu, \\ \Gamma_{\mu\nu}^\alpha &= -\frac{1}{2} g^{\lambda\beta} (\partial_\mu g_{\beta\nu} + \partial_\nu g_{\beta\mu} - \partial_\beta g_{\mu\nu}), \quad F_{\mu\nu} = \partial_\mu b_\nu - \partial_\nu b_\mu. \end{aligned} \quad (4.158)$$

It is a geodesic equation with respect to the affine connection

$$K_\mu^\alpha = \Gamma_{\mu\nu}^\alpha \dot{q}^\nu - \frac{1}{m} g^{\lambda\nu} F_{\mu\nu}$$

of type (4.153). For example, let g be a metric gravitational field and let $b_\mu = eA_\mu$, where A_μ is an electromagnetic potential whose gauge holds fixed. Then the equation (4.158) is the well-known equation of motion of a relativistic massive charge in the presence of these fields.

Let us now perform the quantization of RM, following the standard geometrical quantization of the cotangent bundle (see [Blattner (1983); Sniatycki (1980); Woodhouse (1992)]). As the canonical symplectic form Ω (4.144) on T^*Q is exact, the prequantum bundle is defined as a trivial complex line bundle C over T^*Q . Note that this bundle need no metaplectic correction since T^*X is with canonical coordinates for the symplectic form Ω . Thus, C is called the *quantum bundle*. Let its trivialization

$$C \cong T^*Q \times \mathbb{C} \quad (4.159)$$

hold fixed, and let (t, q^i, p_α, c) , with $c \in \mathbb{C}$, be the associated bundle coordinates. Then one can treat sections of C (4.159) as smooth complex functions on T^*Q . Note that another trivialization of C leads to an equivalent quantization of T^*Q .

Recall that the *Kostant–Souriau prequantization* formula associates to each smooth real function $f \in C^\infty(T^*Q)$ on T^*Q the first-order differential operator

$$\widehat{f} = -i\nabla_{\vartheta_f} + f \quad (4.160)$$

on sections of C , where $\vartheta_f = \partial^\alpha f \partial_\alpha - \partial_\alpha f \partial^\alpha$ is the Hamiltonian vector-field of f and ∇ is the covariant differential with respect to a suitable $U(1)$ -principal connection A on C . This connection preserves the *Hermitian metric* $g(c, c') = \bar{c}c'$ on C , and its curvature form obeys the prequantization condition $R = i\Omega$. For the sake of simplicity, let us assume that Q and, consequently, T^*Q is simply-connected. Then the connection A up to gauge transformations is

$$A = dp_\alpha \otimes \partial^\alpha + dq^\alpha \otimes (\partial_\alpha + icp_\alpha \partial_c), \quad (4.161)$$

and the prequantization operators (4.160) read

$$\widehat{f} = -i\vartheta_f + (f - p_\alpha \partial^\alpha f). \quad (4.162)$$

Let us choose the vertical polarization on T^*Q . It is the vertical tangent bundle VT^*Q of the fibration $\pi : T^*Q \rightarrow Q$. As was mentioned above, the corresponding quantum algebra $\mathcal{A} \subset C^\infty(T^*Q)$ consists of affine functions f (4.147) of momenta p_α . Its representation by operators (4.162) is defined in the space E of sections ρ of the quantum bundle C of compact support which obey the condition $\nabla_{\vartheta\rho} = 0$ for any vertical Hamiltonian vector-field ϑ on T^*Q . This condition takes the form

$$\partial_\alpha f \partial^\alpha \rho = 0, \quad (f \in C^\infty(Q)).$$

It follows that elements of E are independent of momenta and, consequently, fail to be compactly supported, unless $\rho = 0$. This is the well-known problem of Schrödinger quantization which is solved as follows [Blatner (1983); Giachetta *et. al.* (2002b)].

Let $i_Q : Q \rightarrow T^*Q$ be the canonical zero section of the cotangent bundle T^*Q . Let $C_Q = i_Q^*C$ be the pull-back of the bundle C (4.159) over Q . It is a trivial complex line bundle $C_Q = Q \times \mathbb{C}$ provided with the pull-back Hermitian metric $g(c, c') = \bar{c}c'$ and the pull-back

$$A_Q = i_Q^*A = dq^\alpha \otimes (\partial_\alpha + icp_\alpha \partial_c)$$

of the connection A (4.161) on C . Sections of C_Q are smooth complex functions on Q , but this bundle need metaplectic correction.

Let the cohomology group $H^2(Q; \mathbb{Z}_2)$ of Q be trivial. Then a metilinear bundle \mathcal{D} of complex half-forms on Q is defined. It admits the canonical lift of any vector-field τ on Q such that the corresponding Lie derivative of its sections reads

$$\mathfrak{L}_\tau = \tau^\alpha \partial_\alpha + \frac{1}{2} \partial_\alpha \tau^\alpha.$$

Let us consider the tensor product $Y = C_Q \otimes \mathcal{D}$ over Q . Since the Hamiltonian vector-fields

$$\vartheta_f = a^\alpha \partial_\alpha - (p_\mu \partial_\alpha a^\mu + \partial_\alpha b) \partial^\alpha$$

of functions f (4.147) are projected onto Q , one can assign to each element f of the quantum algebra \mathcal{A} the first-order differential operator

$$\widehat{f} = (-i\bar{\nabla}_{\pi\vartheta_f} + f) \otimes \text{Id} + \text{Id} \otimes \mathfrak{L}_{\pi\vartheta_f} = -ia^\alpha \partial_\alpha - \frac{i}{2} \partial_\alpha a^\alpha + b$$

on sections ρ_Q of Y . For the sake of simplicity, let us choose a trivial metilinear bundle $\mathcal{D} \rightarrow Q$ associated to the orientation of Q . Its sections can be written in the form $\rho_Q = (-g)^{1/4} \psi$, where ψ are smooth complex functions on Q . Then the quantum algebra \mathcal{A} can be represented by the operators \widehat{f} (4.148) in the space $\mathbb{C}^\infty(Q)$ of these functions. It can be justified that these operators obey the *Dirac condition*

$$[\widehat{f}, \widehat{f'}] = -i\widehat{\{f, f'\}}.$$

One usually considers the subspace $E_Q \subset \mathbb{C}^\infty(Q)$ of functions of compact support. It is a pre-Hilbert space with respect to the non-degenerate

Hermitian form

$$\langle \psi | \psi' \rangle = \int_Q \psi \bar{\psi}' (-g)^{1/2} d^{m+1}q$$

Note that \hat{f} (4.148) are symmetric operators $\hat{f} = \hat{f}^*$ in E_Q , i.e., $\langle \hat{f}\psi | \psi' \rangle = \langle \psi | \hat{f}\psi' \rangle$. However, the space E_Q gets no physical meaning in RM.

As was mentioned above, the function H_T (4.145) need not belong to the quantum algebra \mathcal{A} , but a polynomial function H_T can be quantized as an element of the enveloping algebra $\bar{\mathcal{A}}$ by operators \hat{H}_T (4.150). Then the quantum constraint (4.146) serves as a relativistic quantum equation.

Let us again consider a massive relativistic charge whose relativistic Hamiltonian is

$$H = \frac{1}{2m} g^{\mu\nu} (p_\mu - eA_\mu)(p_\nu - eA_\nu).$$

It defines the constraint

$$H_T = \frac{1}{m^2} g^{\mu\nu} (p_\mu - eA_\mu)(p_\nu - eA_\nu) - 1 = 0. \quad (4.163)$$

Let us represent the function H_T (4.163) as symmetric product of affine functions of momenta,

$$H_T = \frac{(-g)^{-1/4}}{m} (p_\mu - eA_\mu) (-g)^{1/4} g^{\mu\nu} (-g)^{1/4} (p_\nu - eA_\nu) \frac{(-g)^{-1/4}}{m} - 1.$$

It is quantized by the rule (4.150), where

$$(-g)^{1/4} \circ \hat{\partial}_\alpha \circ (-g)^{-1/4} = -i\partial_\alpha.$$

Then the well-known relativistic quantum equation

$$(-g)^{-1/2} [(\partial_\mu - ieA_\mu) g^{\mu\nu} (-g)^{1/2} (\partial_\nu - ieA_\nu) + m^2] \psi = 0. \quad (4.164)$$

is reproduced up to the factor $(-g)^{-1/2}$.

4.12 Symplectic Structures on Fiber Bundles

In this section, following [Lalonde *et al.* (1998); Lalonde *et al.* (1999); Lalonde and McDuff (2002)], we analyze general symplectic structures on fiber bundles. We first discuss how to characterize *Hamiltonian bundles* and their automorphisms, and then describe their main properties, in particular deriving conditions under which the cohomology of the total space splits as

a product. Finally we state some applications to the action of $Ham(M)$ on M and to non-Hamiltonian symplectic bundles.

4.12.1 Hamiltonian Bundles

4.12.1.1 Characterizing Hamiltonian Bundles

Recall that a fiber bundle $M \rightarrow P \rightarrow B$ is said to be *symplectic fibre bundle* if its *structural group* reduces to the *group of symplectomorphisms* $Symp(M, \omega)$ of the closed symplectic manifold (M, ω) . In this case, each fiber $M_b = \pi^{-1}(b)$ is equipped with a well defined symplectic form ω_b such that (M_b, ω_b) is symplectomorphic to (M, ω) . Our first group of results establish geometric criteria for a symplectic bundle to be Hamiltonian, i.e., for the structural group to reduce to $Ham(M, \omega)$. Quite often we simplify the notation by writing $Ham(M)$ and $Symp_0(M)$ (or even Ham and $Symp_0$) instead of $Ham(M, \omega)$ and $Symp_0(M, \omega)$ (see [Lalonde *et al.* (1998); Lalonde *et al.* (1999); Lalonde and McDuff (2002)]).

Recall that the group $Ham(M, \omega)$ is a connected normal subgroup of the identity component $Symp_0(M, \omega)$ of the group of symplectomorphisms, and fits into the exact sequence

$$\{\text{Id}\} \rightarrow Ham(M, \omega) \rightarrow Symp_0(M, \omega) \xrightarrow{\text{Flux}} H^1(M, \mathbb{R})/\Gamma_\omega \rightarrow \{0\},$$

where Γ_ω is the *flux group*. As $Ham(M)$ is connected, *every Hamiltonian bundle is symplectically trivial over the 1-skeleton of the base*. The following proposition was proved in [McDuff and Salamon (1998)]: A symplectic bundle $\pi : P \rightarrow B$ is Hamiltonian iff the following conditions hold:

- (i) the restriction of π to the 1-skeleton B_1 of B is symplectically trivial, and
- (ii) there is a cohomology class $a \in H^2(P, \mathbb{R})$ that restricts to $[\omega_b]$ on M_b .

There is no loss of generality in assuming that the bundle $\pi : P \rightarrow B$ is smooth. Then recall from [Guillemin *et al.* (1998)] that any 2-form τ on P that restricts to ω_b on each fiber M_b defines a connection ∇_τ on P whose horizontal distribution Hor_τ is just the τ -orthogonal complement of the tangent spaces to the fibers:

$$Hor_\tau(x) = \{v \in T_x P : \tau(v, w) = 0 \text{ for all } w \in T_x(M_{\pi(x)})\}.$$

Such forms τ are called *connection forms*. The closedness of τ is a sufficient (but not necessary) condition for the holonomy of ∇_τ to be sym-

plectic. A simple argument due to [McDuff and Salamon (1998)] shows that the cohomological condition (ii) above is equivalent to the existence of a closed extension τ of the forms ω_b . Condition (i) is then equivalent to requiring that the holonomy of ∇_τ around any loop in B belongs to the identity component $Symp_0(M)$ of $Symp(M)$. Hence the above result can be rephrased in terms of such closed extensions τ as follows: A symplectic bundle $\pi : P \rightarrow B$ is Hamiltonian iff the forms ω_b on the fibers have a closed extension τ such that the holonomy of ∇_τ around any loop in B lies in the identity component $Symp_0(M)$ of $Symp(M)$.

This is a slight extension of a result of [Guillemin *et al.* (1998)], who called a specific choice of τ the *coupling form*. As we show below, the existence of τ is the key to the good behavior of Hamiltonian bundles under composition.

When M is simply connected, $Ham(M)$ is the identity component $Symp_0(M)$ of $Symp(M)$, and so a symplectic bundle is Hamiltonian iff condition (i) above is satisfied, i.e., iff it is trivial over the 1-skeleton B_1 . In this case, as observed by [Gotay *et al.* (1983)], it is known that (i) implies (ii) for general topological reasons to do with the behavior of evaluation maps. More generally, (i) implies (ii) for all symplectic bundles with fiber (M, ω) iff the flux group Γ_ω vanishes.

4.12.1.2 Hamiltonian Structures

The question then arises as to what a *Hamiltonian structure* on a fiber bundle actually is [Lalonde *et al.* (1998); Lalonde *et al.* (1999); Lalonde and McDuff (2002)]. That is, how many Hamiltonian structures can one put on a given *symplectic bundle* $\pi : P \rightarrow B$? And, what does one mean by an automorphism of such a structure?

In homotopy theoretic terms, a Hamiltonian structure on a symplectic bundle $\pi : P \rightarrow B$ is simply a lift \tilde{g} to $BHam(M)$ of the classifying map $g : B \rightarrow BSymp(M, \omega)$ of the underlying symplectic bundle, i.e., it is a homotopy commutative diagram

$$\begin{array}{ccc}
 & B\operatorname{Ham}(M) & \\
 \tilde{g} \nearrow & \downarrow 1 & \\
 B & \xrightarrow{g} & B\operatorname{Symp}(M)
 \end{array}$$

Hamiltonian structures are in bijective correspondence with homotopy classes of such lifts. There are two stages to choosing the lift \tilde{g} : one first lifts g to a map \hat{g} into $B\operatorname{Symp}_0(M, \omega)$, where Symp_0 is the identity component of Symp , and then to a map \tilde{g} into $B\operatorname{Ham}(M, \omega)$. As we will show below, choosing \hat{g} is equivalent to fixing the isotopy class of an identification of (M, ω) with the fiber (M_{b_0}, ω_{b_0}) over the base point b_0 . If B is simply connected, in particular if B is a single point, there is then a unique Hamiltonian structure on P , i.e., a unique choice of lift \tilde{g} . Before describing what happens in the general case, we discuss properties of the extensions τ .

Let $\tau \in \Omega^2(P)$ be a closed extension of the symplectic forms on the fibers. Given a loop, $\gamma : S^1 \rightarrow B$, based at b_0 , and a symplectic trivialization $T_\gamma : \gamma^*(P) \rightarrow S^1 \times (M, \omega)$ that extends the given identification of M_{b_0} with M , push forward τ to a form $(T_\gamma)_*\tau$ on $S^1 \times (M, \omega)$. Its characteristic flow round S^1 is transverse to the fibers and defines a symplectic isotopy ϕ_t of $(M, \omega) = (M_{b_0}, \omega_{b_0})$ whose flux, as a map from $H_1(M) \rightarrow \mathbb{R}$, is equal to $(T_\gamma)_*[\tau]([S^1] \otimes \cdot)$. This flux depends only on the cohomology class a of τ . Moreover, as we mentioned above, any extension a of the fiber class $[\omega]$ can be represented by a form τ that extends the ω_b . Thus, given T_γ and an extension $a = [\tau] \in H^2(P)$ of the fiber symplectic class $[\omega]$, it makes sense to define the *flux class* $f(T_\gamma, a) \in H^1(M, \mathbb{R})$ by

$$f(T_\gamma, a)(\delta) = (T_\gamma)_*(a)(\gamma \otimes \delta) \quad \text{for all } \delta \in H_1(M).$$

The equivalence class $[f(T_\gamma, a)] \in H^1(M, \mathbb{R})/\Gamma_\omega$ does not depend on the choice of T_γ : indeed two such choices differ by a loop ϕ in $\operatorname{Symp}_0(M, \omega)$ and so the difference

$$f(T_\gamma, a) - f(T'_\gamma, a) = f(T_\gamma, a) \circ \operatorname{Tr}_\phi = \omega \circ \operatorname{Tr}_\phi$$

belongs to Γ_ω . The following lemma is elementary: If $\pi : P \rightarrow B$ is a symplectic bundle satisfying the above conditions, there is an extension a

of the symplectic fiber class that has a *trivial flux*

$$[f(T_\gamma, a)] = 0 \in H^1(M, \mathbb{R})/\Gamma_\omega \quad \text{around each loop } \gamma \in B.$$

An extension a of the symplectic fiber class $[\omega_{b_0}]$ is normalized if it satisfies the conclusions of the above lemma. Two such extensions a and a' are equivalent (in symbols, $a \sim a'$) iff they have equal restrictions to $\pi^{-1}(B_1)$, or, equivalently, iff $a - a' \in \pi^*(H^2(B))$.

We show below that Hamiltonian structures are in one-to-one correspondence with symplectic trivializations of the 1-skeleton B_1 of B , with two such trivializations being equivalent iff they differ by Hamiltonian loops. If two trivializations T_γ, T'_γ differ by a Hamiltonian loop ϕ then $f(T_\gamma, a) - f(T'_\gamma, a) = 0$. In terms of fluxes of closed extensions, we therefore get the following theorem: Assume that a symplectic bundle $\pi : P \rightarrow B$ can be symplectically trivialized over B_1 . Then a Hamiltonian structure exists on P iff there is a normalized extension a of ω . Such a structure consists of an isotopy class of symplectomorphisms $(M, \omega) \rightarrow (M_{b_0}, \omega_{b_0})$ together with an equivalence class $\{a\}$ of normalized extensions of the fiber symplectic class.

In other words, with respect to a fixed trivialization over B_1 , Hamiltonian structures are in one-to-one correspondence with homomorphisms $\pi_1(B) \rightarrow \Gamma_\omega$, given by the fluxes $f_\gamma(T, a)$ of monodromies round the loops of the base. We will call $\{a\}$ the *Hamiltonian extension class*, and will denote the Hamiltonian structure on P by the triple $(P, \pi, \{a\})$.

We now turn to the question of describing automorphisms of Hamiltonian structures [Lalonde *et al.* (1998); Lalonde *et al.* (1999)]. It is convenient to distinguish between symplectic and Hamiltonian automorphisms, just as we distinguish between $Symp(M, \omega)$ and $Ham(M, \omega)$ in the case when $B = pt$. Notice that if $P \rightarrow B$ is a symplectic bundle, there is a natural notion of symplectic automorphism. This is a fiberwise diffeomorphism $\Phi : P \rightarrow P$ that covers the identity map on the base and restricts on each fiber to an element Φ_b of the group $Symp(M_b, \omega_b)$. Because $Ham(M, \omega)$ is a normal subgroup of $Symp(M, \omega)$, it also makes sense to require that $\Phi_b \in Ham(M_b, \omega_b)$ for each b . Such automorphisms are called *Hamiltonian automorphisms* of the symplectic bundle $P \rightarrow B$. Let us write $Symp(P, \pi)$ and $Ham(P, \pi)$ for the groups of such automorphisms. Observe that the group $Ham(P, \pi)$ may not be connected. Because the fibers of Hamiltonian bundles are identified with (M, ω) up to isotopy, we shall also need to consider the (not necessarily connected) group $Symp_0(P, \pi)$ of symplectomorphisms of (P, π) where $\Phi_b \in Symp_0(M_b, \omega_b)$ for one and hence all

b.

Let us consider automorphisms of Hamiltonian bundles [Lalonde and McDuff (2002)]. As a guide note that in the trivial case when $B = pt$, a Hamiltonian structure on P is an identification of P with M up to symplectic isotopy. Hence the group of automorphisms of this structure can be identified with $Symp_0(M, \omega)$. In general, if $\{a\}$ is a Hamiltonian structure on (P, π) and $\Phi \in Symp_0(P, \pi)$ then $\Phi^*(\{a\}) = \{a\}$ iff $\Phi^*(a) = a$ for some a in the class $\{a\}$, because Φ induces the identity map on the base and $a - a' \in \pi^*(H^2(B))$ when $a \sim a'$. We therefore make the following definition.

Let $(P, \pi, \{a\})$ be a Hamiltonian structure on the symplectic bundle $P \rightarrow B$ and let $\Phi \in Symp(P, \pi)$. Then Φ is an *automorphism of the Hamiltonian structure* $(P, \pi, \{a\})$ if $\Phi \in Symp_0(P, \pi)$ and $\Phi^*(\{a\}) = \{a\}$. The group formed by these elements is denoted by $Aut(P, \pi, \{a\})$.

The following result is not hard to prove, but is easiest to see in the context of a discussion of the action of $Ham(M)$ on $H^*(M)$.

Let $P \rightarrow B$ be a Hamiltonian bundle and $\Phi \in Symp_0(P, \pi)$. Then the following statements are equivalent [Lalonde and McDuff (2002)]:

- (i) Φ is isotopic to an element of $Ham(P, \pi)$;
- (ii) $\Phi^*(\{a\}) = \{a\}$ for some Hamiltonian structure $\{a\}$ on P ;
- (iii) $\Phi^*(\{a\}) = \{a\}$ for all Hamiltonian structures $\{a\}$ on P .

For any Hamiltonian bundle $P \rightarrow B$, the group $Aut(P, \pi, \{a\})$ does not depend on the choice of the Hamiltonian structure $\{a\}$ put on P . Moreover, it contains $Ham(P, \pi)$ and each element of $Aut(P, \pi, \{a\})$ is isotopic to an element in $Ham(P, \pi)$.

The following characterization is now obvious:

Let P be the product $B \times M$ and $\{a\}$ any Hamiltonian structure. Then:

- (i) $Ham(P, \pi)$ consists of all maps from B to $Ham(M, \omega)$.
- (ii) $Aut(P, \pi, \{a\})$ consists of all maps $\Phi : B \rightarrow Symp_0(M, \omega)$ for which the following composition is trivial

$$\pi_1(B) \xrightarrow{\Phi_*} \pi_1(Symp_0(M)) \xrightarrow{Flux_\omega} H^1(M, \mathbb{R}).$$

The basic reason why the above proposition holds is that Hamiltonian automorphisms of (P, π) act trivially on the set of extensions of the fiber symplectic class. This need not be true for symplectic automorphisms. For example, if

$$\pi : P = S^1 \times M \rightarrow S^1$$

is a trivial bundle and Φ is given by a non-Hamiltonian loop ϕ in $Symp_0(M)$, then Φ is in $Symp_0(P, \pi)$ but it preserves no Hamiltonian structure on P since

$$\Phi^*(a) = a + [dt] \otimes Flux(\phi).$$

In general, if we choose a trivialization of P over B_1 , there are exact sequences

$$\begin{aligned} \{\text{Id}\} &\rightarrow \text{Aut}(P, \pi, \{a\}) \rightarrow \text{Symp}_0(P, \pi) \rightarrow \text{Hom}(\pi_1(B), \Gamma_\omega) \rightarrow \{\text{id}\}, \\ \{\text{Id}\} &\rightarrow \text{Ham}(P, \pi, \{a\}) \rightarrow \text{Aut}(P, \pi, \{a\}) \rightarrow H^1(M, \mathbb{R})/\Gamma_\omega \rightarrow \{0\}. \end{aligned}$$

In particular, the subgroup of $\text{Aut}(P, \pi, \{a\})$ consisting of automorphisms that belong to $\text{Ham}(M_{b_0}, \omega_{b_0})$ at the base point b_0 retracts to $\text{Ham}(P, \pi, \{a\})$.

4.12.1.3 Marked Hamiltonian Structures

Another approach to characterizing a Hamiltonian structure is to define it in terms of a structure on the fiber that is preserved by elements of the Hamiltonian group [Lalonde and McDuff (2002)].

The so-called *marked symplectic manifold* $(M, \omega, [L])$ is a pair consisting of a closed symplectic manifold (M, ω) together with a marking $[L]$. Here L is a collection $\{\ell_1, \dots, \ell_k\}$ of loops $\ell_i : S^1 \rightarrow M$ in M that projects to a minimal generating set $\mathcal{G}_L = \{[\ell_1], \dots, [\ell_k]\}$ for $H_1(M, \mathbb{Z})/\text{torsion}$. A marking $[L]$ is an equivalence class of generating loops L , where $L \sim L'$ if for each i there is an singular integral 2-chain c_i whose boundary modulo torsion is $\ell'_i - \ell_i$ such that $\int_{c_i} \omega = 0$.

The symplectomorphism group acts on the space \mathcal{L} of markings. Moreover, it is easy to check that if a symplectomorphism ϕ fixes one marking $[L]$ it fixes them all. Hence the group

$$L\text{Ham}(M, \omega) = L\text{Ham}(M, \omega, [L]) = \{\phi \in \text{Symp}(M, \omega) : \phi_*[L] = [L]\}$$

independent of the choice of $[L]$. Its identity component is $\text{Ham}(M, \omega)$.

There is a forgetful map $[L] \rightarrow \mathcal{G}_L$ from the space \mathcal{L} of markings to the space of minimal generating sets for the group $H_1(M, \mathbb{Z})$, and it is not hard to check that its fiber is $(\mathbb{R}/\mathcal{P})^k$, where \mathcal{P} is the image of the period homomorphism

$$I_{[\omega]} : H_2(M, \mathbb{Z}) \rightarrow \mathbb{R}.$$

If \mathcal{P} is not discrete, there is no nice topology one can put on \mathcal{L} . However, it has a pseudotopology, i.e., one can specify which maps of finite polyhedra X into \mathcal{L} are continuous, namely: $f : X \rightarrow \mathcal{L}$ is continuous iff every $x \in X$ has a neighborhood U_x such that $f : U_x \rightarrow \mathcal{L}$ lifts to a continuous map into the space of generating loops L .

Now, let us fix a marking $[L]$ on (M, ω) . A Hamiltonian structure on a symplectic bundle $\pi : P \rightarrow B$ is an isotopy class of symplectomorphisms $(M, \omega, [L]) \rightarrow (M_{b_0}, \omega_{b_0}, [L_0])$ together with a continuous choice of marking $[L_b]$ on each fiber (M_b, ω_b) that is trivial over the 1-skeleton B_1 of B in the sense that there is a symplectic trivialization

$$\Phi : \pi^{-1}(B_1) \rightarrow B_1 \times (M, \omega, [L])$$

that respects the markings on each fiber.

Here is another way of thinking of a Hamiltonian structure due to Polterovich. He observed that there is an exact sequence

$$0 \rightarrow \mathbb{R}/\mathcal{P} \rightarrow SH_1(M, \omega) \rightarrow H_1(M, \mathbb{Z}) \rightarrow 0,$$

where $SH_1(M, \omega)$ is the so-called *strange homology group* formed by quotienting the space of integral 1-cycles by the image under d of the space of integral 2-chains with zero symplectic area. The group $Symp(M, \omega)$ acts on $SH_1(M, \omega)$. Moreover, if $\phi \in Symp_0(M)$ and $\tilde{a} \in SH_1(M)$ projects to $a \in H_1(M)$, then $\phi_*(\tilde{a}) - \tilde{a} \in \mathbb{R}/\mathcal{P}$ can be thought of as the value of the class $Flux(\phi) \in H^1(M, \mathbb{R})/\Gamma_\omega$ on a . It is easy to see that $LHam(M, \omega)$ is the subgroup of $Symp(M, \omega)$ that acts trivially on $SH_1(M, \mathbb{Z})$. Further, a marking on (M, ω) is a pair consisting of a splitting of the above sequence together with a generating set $_L$ for $H_1(M, \mathbb{Z})/\text{torsion}$.

Given any symplectic bundle $P \rightarrow B$ there is an associated bundle of abelian groups with fiber $SH_1(M, \omega)$. A Hamiltonian structure on $P \rightarrow B$ is a flat connection on this bundle that is trivial over the 1-skeleton B_1 , under an appropriate equivalence relation.

These ideas can obviously be generalized to bundles that are not trivial over the 1-skeleton. Equivalently, one can consider bundles with disconnected structural group. This group could be the whole of $LHam(M, \omega)$. One could also restrict to elements acting trivially on $H^*(M)$ and/or to those that act trivially on the groups

$$SH_{2k-1}(M, \omega) = \frac{\text{integral } (2k-1)\text{-cycles}}{d(2k\text{-chains in the kernel of } \omega^k)}.$$

These generalizations of $SH_1(M, \omega)$ are closely connected to Reznikov's Futaki type characters [Reznikov (1997)]. It is not yet clear what is the most natural disconnected extension of $Ham(M, \omega)$.

4.12.1.4 Stability

Another important property of Hamiltonian bundles is stability [Lalonde and McDuff (2002)]. A symplectic (resp. Hamiltonian) bundle $\pi : P \rightarrow B$ with fiber (M, ω) is said to be stable if π may be given a symplectic (resp. Hamiltonian) structure with respect to any symplectic form ω' on M that is sufficiently close to (but not necessarily cohomologous to) ω , in such a way that the structure depends continuously on ω' .

Using *Moser's homotopy argument*, it is easy to prove that any symplectic bundle is stable. The following characterization of Hamiltonian stability is an almost immediate consequence of above theorem:

A Hamiltonian bundle $\pi : P \rightarrow B$ is stable iff the restriction map $H^2(P, \mathbb{R}) \rightarrow H^2(M, \mathbb{R})$ is surjective.

The following result is less immediate: Every Hamiltonian bundle is stable.

The proof uses the (difficult) stability property for Hamiltonian bundles over S^2 that was established in [Lalonde *et al.* (1999); McDuff (2000)] as well as the (easy) fact that the image of the evaluation map $\pi_2(Ham(M)) \rightarrow \pi_2(M)$ lies in the kernel of $[\omega]$.

4.12.1.5 Cohomological Splitting

We next extend the splitting results of [Lalonde *et al.* (1999); McDuff (2000)], which prove that the *rational cohomology of every Hamiltonian bundle* $\pi : P \rightarrow S^2$ *splits additively*, i.e., there is an additive isomorphism

$$H^*(P) \cong H^*(S^2) \otimes H^*(M).$$

For short we will say in this situation that π is *c-split*. This is a deep result, that requires the use of *Gromov–Witten invariants* for its proof.¹⁶ The results of the present subsection provide some answers to the following

¹⁶Recall that Gromov–Witten invariants are rational numbers that count pseudo-holomorphic curves meeting prescribed conditions in a given symplectic manifold. These invariants may be packaged as a homology or cohomology class in an appropriate space, or as the deformed cup product of quantum cohomology. They have been used to distinguish symplectic manifolds that were previously indistinguishable. They also play a crucial role in closed type IIA string theory.

question [Lalonde and McDuff (2002)]: *Does any Hamiltonian fiber bundle over a compact CW-complex c -split?*

A special case is when the structural group of $P \rightarrow B$ can be reduced to a compact Lie subgroup G of $\text{Ham}(M)$. Here c -splitting over any base follows from the work of Atiyah–Bott [Atiyah and Bott (1984)]. In this context, one usually discusses the universal Hamiltonian G -bundle with fiber M

$$M \longrightarrow M_G = EG \times_G M \longrightarrow BG.$$

The cohomology of $P = M_G$ is known as the equivariant cohomology $H_G^*(M)$ of M . Atiyah–Bott show that if G is a torus T that acts in a Hamiltonian way on M then the bundle $M_T \rightarrow BT$ is c -split. The result for a general compact Lie group G follows by standard arguments [Lalonde and McDuff (2002)].

The following theorem describes conditions on the base B that imply c -splitting. Let (M, ω) be a closed symplectic manifold, and $M \hookrightarrow P \rightarrow B$ a bundle with structure group $\text{Ham}(M)$ and with base a compact CW-complex B . Then the rational cohomology of P splits in each of the following cases:

- (i) the base has the homotopy type of a coadjoint orbit or of a product of spheres with at most three of dimension 1;
- (ii) the base has the homotopy type of a complex blow up of a product of complex projective spaces;
- (iii) $\dim(B) \leq 3$.

Case (ii) is a generalization of the foundational example $B = S^2$ and is proved by similar analytic methods. The idea is to show that the map $\iota : H_*(M) \rightarrow H_*(P)$ is injective by showing that the image $\iota(a)$ in P of any class $a \in H_*(M)$ can be detected by a nonzero Gromov–Witten invariant of the form $n_P(\iota(a), c_1, \dots, c_n; \sigma)$, where $c_i \in H_*(P)$ and $\sigma \in H_2(P)$ is a spherical class with nonzero image in $H_2(B)$. The proof should generalize to the case when all one assumes about the base is that there is a nonzero invariant of the form $n_B(pt, pt, c_1, \dots, c_k; A)$ [Lalonde *et al.* (1998); Lalonde *et al.* (1999)].

The proofs of parts (i) and (iii) start from the fact of c -splitting over S^2 and proceed using purely topological methods. The following fact about compositions of Hamiltonian bundles is especially useful. Let $M \hookrightarrow P \rightarrow B$ be a Hamiltonian bundle over a simply connected base B and assume

that all Hamiltonian bundles over M as well as over B c -split. Then any Hamiltonian bundle over P c -splits too.¹⁷ This provides a powerful recursive argument which allows one to establish c -splitting over \mathbf{CP}^n by induction on n , and is an essential tool in all arguments here [Lalonde and McDuff (2002)].

It is natural to wonder whether c -splitting is a purely homotopy-theoretic property. A c -symplectic manifold (M, a_M) is defined to be a $2n$ -manifold together with a class $a_M \in H^2(M)$ such that $a_M^n > 0$. In view of above theorem one could define a c -Hamiltonian bundle over a simply connected base manifold B to be a bundle $P \rightarrow B$ with c -symplectic fiber (M, a_M) in which the symplectic class a_M extends to a class $a \in H^2(P)$. A variety of results about symplectic torus actions were discussed in [All-day (1998)], some of which do extend to the c -symplectic case and some of which do not. The next lemma shows that c -splitting in general is a geometric rather than a homotopy-theoretic property: There is a c -Hamiltonian bundle over S^2 that is not c -split.

It is also worth noting that it is essential to restrict to finite dimensional spaces: c -splitting does not always hold for ‘Hamiltonian’ bundles with infinite dimensional fiber [Lalonde and McDuff (2002)].

4.12.1.6 Homological Action of $\text{Ham}(M)$ on M

The action $\text{Ham}(M) \times M \rightarrow M$ gives rise to mutually dual maps [Lalonde and McDuff (2002)]

$$\begin{aligned} \text{Tr}_\phi : H_k(\text{Ham}(M)) \otimes H_*(M) &\longrightarrow H_{k+*}(M), & (\phi, Z) &\mapsto \text{Tr}_\phi(Z), & \text{and} \\ \text{Tr}_\phi^* : H_k(\text{Ham}(M)) &\longrightarrow \text{Hom}(H^*(M), H^{*-k}(M)), & & (k \geq 0). \end{aligned}$$

In this language, the flux of a loop $\phi \in \pi_1(\text{Ham}(M))$ is precisely the element $\text{Tr}_\phi^*([\omega]) \in H^1(M)$. (Here we should use real rather than rational coefficients so that $[\omega] \in H^*(M)$.) The following result is a consequence of above theorem: The maps Tr_ϕ and Tr_ϕ^* are zero for all $\phi \in H_k(\text{Ham}(M))$, $k > 0$.

The argument goes as follows. Recall that the cohomology ring of $\text{Ham}(M)$ is generated by elements dual to its homotopy. It therefore suffices to consider the restriction of Tr_k to the spherical elements ϕ . But in this case it is not hard to see that the Tr_k are precisely the connecting homomorphisms in the Wang sequence of the bundle $P_\phi \rightarrow S^{k+1}$ with clutching function ϕ . These vanish because all Hamiltonian bundles over

¹⁷This fact is based on the characterization of Hamiltonian bundles in terms of closed extensions of the symplectic form.

spheres are c -split [Lalonde and McDuff (2002)].

In particular, looking at the action on $H_0(M)$, we see that the *point evaluation map*

$$\mathfrak{ev} : \text{Ham}(M) \rightarrow M : \quad \psi \mapsto \psi(x)$$

induces the trivial map on rational (co)homology. It also induces the trivial map on π_1 . However, the map on $\pi_k, k > 1$, need not be trivial. To see this, consider the action of $\text{Ham}(M)$ on the symplectic frame bundle $SFr(M)$ of M and the corresponding point evaluation maps. The obvious action of $SO(3) \simeq \text{Ham}(S^2)$ on $SFr(S^2) \simeq \mathbf{RP}^3$ induces an isomorphism

$$H_3(SO(3)) \cong H_3(SFr(S^2)),$$

showing that these evaluation maps are not homologically trivial. Moreover, its composite with the projection $SFr(S^2) \rightarrow S^2$ gives rise to a nonzero map

$$\pi_3(SO(3)) = \pi_3(\text{Ham}(S^2)) \rightarrow \pi_3(S^2).$$

Thus the corresponding Hamiltonian fibration over S^4 with fiber S^2 , though c -split, does not have a section.

Note, however, that the extended evaluation map

$$\pi_{2\ell}(X^X) \xrightarrow{\mathfrak{ev}} \pi_{2\ell}(X) \rightarrow H_{2\ell}(X, \mathbf{Q}), \quad \ell > 0,$$

is always zero, if X is a finite CW complex and X^X is its space of self-maps. Indeed, because the cohomology ring $H^*(X^X, \mathbf{Q})$ is freely generated by elements dual to $\pi_*(X^X) \otimes \mathbf{Q}$, there would otherwise be an element $a \in H^{2\ell}(X)$ that would pull back to an element of infinite order in the cohomology ring of the H -space X^X . Hence a itself would have to have infinite order, which is impossible. A more delicate argument shows that the integral evaluation $\pi_{2\ell}(X^X) \rightarrow H_{2\ell}(X, \mathbb{Z})$ is zero [Gottlieb (1975)].

A Hamiltonian automorphism of the product Hamiltonian bundle $B \times M \rightarrow B$ is simply a map $B \rightarrow B \times \text{Ham}(M)$ of the form $b \mapsto (b, \phi_b)$. If B is a closed manifold we will see that any Hamiltonian automorphism of the product bundle acts as the identity map on the rational cohomology of $B \times M$. The natural generalization of this result would claim that a Hamiltonian automorphism of a bundle P acts as the identity map on the rational cohomology of P . We do not know yet whether this is true in general. However, we can show that it is closely related to the c -splitting

of Hamiltonian bundles. Thus we can establish it only under conditions similar to the conditions under which c -splitting holds.

4.12.1.7 General Symplectic Bundles

Consider the Wang sequence for a symplectic bundle $\pi : P \rightarrow S^2$ with clutching map $\phi \in \pi_1(\text{Symp}(M))$:

$$\dots \longrightarrow H^k(M) \xrightarrow{\partial} H^{k-1}(M) \xrightarrow{u} H^{k+1}(P) \xrightarrow{restr} H^{k+1}(M) \longrightarrow \dots$$

Here the map u may be realized in de Rham cohomology by choosing any extension of a given closed form α on M and then wedging it with the pullback of a normalized area form on the base. Further, as pointed out above, the boundary map $\partial = \partial_\phi$ is just Tr_ϕ^* . Therefore, the bundle is Hamiltonian iff

$$\text{Tr}_\phi^*([\omega]) = \partial([\omega]) = 0.$$

In the Hamiltonian case the above splitting theorem implies that ∂ is identically 0. In the general case, we know that the map $\partial : H^*(M) \rightarrow H^{*-1}(M)$ is a derivation, i.e.,

$$\partial(ab) = \partial(a)b + (-1)^{\deg(a)}a\partial(b).$$

The following result is an easy consequence of the fact that the action of

$$\pi_2(\text{Ham}(M)) = \pi_2(\text{Symp}(M)) \quad \text{on} \quad H^*(M)$$

is trivial: The boundary map ∂ in the Wang rational cohomology sequence of a symplectic bundle over S^2 has the basic property: $\partial \circ \partial = 0$.

The above result holds trivially when ϕ corresponds to a smooth (not necessarily symplectic) S^1 -action since then ∂ is given in the *de Rham cohomology* by contraction ι_X by the generating vector field X . Moreover, the authors know of no smooth bundle over S^2 for which the above proposition does not hold, though it is likely that they exist. Such a bundle would have no extension over \mathbf{CP}^2 .

One consequence is the following result about the boundary map $\partial = \partial_\phi$ in the case when the loop ϕ is far from being Hamiltonian. Recall (e.g., from [Lalonde *et al.* (1999)]) that $\pi_1(\text{Ham}(M))$ is included in (but not necessarily equal to) the kernel of the evaluation map $\pi_1(\text{Symp}(M)) \rightarrow \pi_1(M)$. Any loop whose evaluation is homologically essential can therefore

be thought of as ‘very non-Hamiltonian’. We have the following corollary:

$$\text{Ker } \partial = \text{Im } \partial$$

iff the image of ϕ under the evaluation map $\pi_1(\text{Symp}(M)) \rightarrow H_1(M, \mathbf{Q})$ is nonzero. A similar result was obtained by [Allday (1998)] concerning S^1 actions on c -symplectic manifolds. He was considering manifolds M that satisfy the weak *Lefschetz condition*, i.e., manifolds such that

$$\wedge[\omega]^{n-1} : H^1(M, \mathbb{R}) \rightarrow H^{2n-1}(M, \mathbb{R})$$

is an isomorphism, in which case every non-Hamiltonian loop is ‘very non-Hamiltonian’.

4.12.1.8 Existence of Hamiltonian Structures

Geometric proofs (such as those in [McDuff and Salamon (1998)]) apply when P and B are smooth manifolds and π is a smooth surjection. However, as the following lemma makes clear, this is no restriction.

Suppose that $\pi : Q \rightarrow W$ is a locally trivial bundle over a finite CW complex W with compact fiber (M, ω) and suppose that the structural group G is equal either to $\text{Symp}(M, \omega)$ or to $\text{Ham}(M, \omega)$. Then there is a smooth bundle $\pi : P \rightarrow B$ as above with structural group G and a homeomorphism f of W onto a closed subset of B such that $\pi : Q \rightarrow W$ is homeomorphic to the pullback bundle $f^*(P) \rightarrow W$.

Let us embed W into some Euclidean space and let B be a suitable small neighborhood of W . Then W is a retract of B so that the classifying map $W \rightarrow BG$ extends to B . It remains to approximate this map $B \rightarrow BG$ by a smooth map. To get a relation between the existence of the class a and the structural group it seems necessary to use the idea of a symplectic connection. We begin with an easy lemma.

Let $P \rightarrow B$ be a symplectic bundle with closed connection form τ . Then the holonomy of the corresponding connection ∇_τ round any contractible loop in B is Hamiltonian.

To prove it, suffices to consider the case when $B = D^2$. Then the bundle $\pi : P \rightarrow D^2$ is symplectically trivial and so may be identified with the product $D^2 \times M$ in such a way that the symplectic form on each fiber is simply ω . Use this trivialization to identify the holonomy round the loop $s \mapsto e^{2\pi is} \in \partial D^2$ with a family of diffeomorphisms $\Phi_s : M \rightarrow M, s \in [0, 1]$. Since this holonomy is simply the flow along the null directions (or characteristics) of the closed form τ on the hypersurface ∂P , a standard

calculation shows that the Φ_s are symplectomorphisms. Given a 1-cycle $\delta : S^1 \rightarrow M$ in the fiber M over $1 \in \partial D^2$, consider the closed 2-cycle that is the union of the following two cylinders:

$$\begin{aligned} c_1 : [0, 1] \times S^1 &\rightarrow \partial D^2 \times M : (s, t) \mapsto (e^{2\pi i s}, \Phi_s(\delta(t))), \\ c_2 : [0, 1] \times S^1 &\rightarrow 1 \times M : (s, t) \mapsto (1, \Phi_{1-s}(\delta(t))). \end{aligned}$$

This cycle is obviously contractible. Hence,

$$\tau(c_1) = -\tau(c_2) = \text{Flux}(\{\Phi_s\})(\delta).$$

But $\tau(c_1) = 0$ since the characteristics of $\tau|_{\partial P}$ are tangent to c_1 . Applying this to all δ , we see that the holonomy round ∂D^2 has zero flux and so is Hamiltonian.

If $\pi_1(B) = 0$ then a symplectic bundle $\pi : P \rightarrow B$ is Hamiltonian iff the class $[\omega_b] \in H^2(M)$ extends to $a \in H^*(P)$.

Suppose first that the class a exists. We can work in the smooth category. Then Thurston's convexity argument allows us to construct a closed connection form τ on P and hence a horizontal distribution Hor_τ . The previous lemma shows that the holonomy around every contractible loop in B is Hamiltonian. Since B is simply connected, the holonomy round *all* loops is Hamiltonian. Using this, it is easy to reduce the structural group of $P \rightarrow B$ to $\text{Ham}(M)$. For more details, see [McDuff and Salamon (1998)].

Next, suppose that the bundle is Hamiltonian. We need to show that the fiber symplectic class extends to P . The proof in [McDuff and Salamon (1998)] does this by the method of [Guillemin *et. al.* (1998)] and constructs a closed connection form τ , called the *coupling form*, starting from a connection with Hamiltonian holonomy. This construction uses the curvature of the connection and is quite analytic. In contrast, we shall now use topological arguments to reduce to the cases $B = S^2$ and $B = S^3$. These cases are then dealt with by elementary arguments.

Consider the *Leray–Serre cohomology spectral sequence* for $M \rightarrow P \rightarrow B$. Its E_2 term is a product: $E_2^{p,q} = H^p(B) \otimes H^q(M)$.¹⁸ We need to show that the class $[\omega] \in E_2^{0,2}$ survives into the E_∞ term, which happens iff it is in the kernel of the two differentials $d_2^{0,2}, d_3^{0,2}$. Now

$$d_2^{0,2} : H^2(M) \rightarrow H^2(B) \otimes H^1(M)$$

is essentially the same as the flux homomorphism. More precisely, if $c : S^2 \rightarrow B$ represents some element (also called c) in $H_2(B)$, then the pullback

¹⁸Here H^* denotes cohomology over \mathbb{R} .

of the bundle $\pi : P \rightarrow B$ by c is a bundle over S^2 that is determined by a loop $\phi_c \in \pi_1(\text{Ham}(M))$ that is well defined up to conjugacy. Moreover, for each $\lambda \in H_1(M)$,

$$d_2^{0,2}([\omega])(c, \lambda) = \text{Tr}_{\phi_c}^*(\lambda),$$

where Tr^* is as above. Hence $d_2^{0,2}([\omega]) = 0$ because ϕ_c is Hamiltonian [Lalonde and McDuff (2002)].

To deal with d_3 observe first that because the inclusion of the 3-skeleton B_3 into B induces an injection $H^q(B) \rightarrow H^q(B_3)$ for $q \leq 3$, $d_3^{0,2}([\omega])$ vanishes in the spectral sequence for $P \rightarrow B$ if it vanishes for the pullback bundle over B_3 . Therefore we may suppose that B is a 3-dimensional CW-complex whose 2-skeleton B_2 is a wedge of 2 spheres, as $\pi_1(B) = 0$. Further, we can choose the cell decomposition so that the first k 3-cells span the kernel of the boundary map $C_3 \rightarrow C_2$ in the cellular chain complex of B_3 . Because $H_2(B_2) = \pi_2(B_2)$, the attaching maps of these first k -cells are null homotopic. Hence there is a wedge B' of 2-spheres and 3-spheres and a map $B' \rightarrow B_3$ that induces a surjection on H_3 . It therefore suffices to show that $d_3^{0,2}([\omega])$ vanishes in the pullback bundle over B' . This will clearly be the case if it vanishes in every Hamiltonian bundle over S^3 .

Now, a Hamiltonian fiber bundle over S^3 is determined by a map

$$I^2/\partial I^2 = S^2 \rightarrow \text{Ham}(M) : (s, t) \mapsto \phi_{s,t},$$

and it is easy to see that $d_3^{0,2}([\omega]) = 0$ exactly when the the evaluation map

$$ev_x : \text{Ham}(M) \rightarrow M : \phi \mapsto \phi(x)$$

takes $\pi_2(\text{Ham}(M))$ into the kernel of ω .

Given a smooth map $\Psi : (I^2, \partial I^2) \rightarrow (\text{Ham}(M), \text{Id})$ and $x \in M$, let $\Psi^x : (I^2, \partial I^2) \rightarrow M$ be the composite of Ψ with evaluation at x . Then we have

$$\int_{I^2} (\Psi^x)^* \omega = 0, \quad \text{for all } x \in M.$$

For each s, t let $X_{s,t}$ (resp. $Y_{s,t}$) be the Hamiltonian vector field on M that is tangent to the flow of the isotopy $s \mapsto \Psi^x(s, t)$, (resp. $t \mapsto \Psi^x(s, t)$.) Then

$$\int_{I^2} (\Psi^x)^* \omega = \iint \omega(X_{s,t}(\Psi^x(s, t)), Y_{s,t}(\Psi^x(s, t))) ds dt.$$

The first observation is that this integral is a constant c that is independent of x , since the maps $\Psi^x : S^2 \rightarrow M$ are all homotopic. Secondly, recall that for any Hamiltonian vector fields X, Y on M

$$\int_M \omega(X, Y) \omega^n = n \int_M \omega(X, \cdot) \omega(Y, \cdot) \omega^{n-1} = 0,$$

since $\omega(X, \cdot), \omega(Y, \cdot)$ are exact 1-forms. Taking $X_{s,t} = X_{s,t}(\Psi^x(s, t))$ and similarly for Y , we have

$$\int c \omega^n = \int_{I^2} \left(\int_M \omega(X_{s,t}, Y_{s,t}) \omega^n \right) ds dt = 0.$$

Hence $c = 0$. This lemma can also be proved by purely topological methods [Lalonde and McDuff (2002)].

Suppose that $\pi_B : P \rightarrow B$ is Hamiltonian. It is classified by a map $B \rightarrow B\text{Ham}(M)$. Because $B\text{Ham}(M)$ is simply connected this factors through a map $C \rightarrow B\text{Ham}(M)$, where C is obtained by collapsing the 1-skeleton of B to a point. In particular condition (i) is satisfied. To verify (ii), let $\pi_C : Q \rightarrow C$ be the corresponding Hamiltonian bundle, so that there is a commutative diagram

$$\begin{array}{ccc} P & \xrightarrow{\quad} & Q \\ \pi_B \downarrow & & \downarrow \pi_C \\ B & \xrightarrow{\quad} & C = B/B_1 \end{array}$$

There is a class $a_C \in H^2(Q)$ that restricts to $[\omega]$ on the fibers. Its pullback to P is the desired class a .

Conversely, suppose that conditions (i) and (ii) are satisfied. By (i), the classifying map $B \rightarrow B\text{Symp}(M)$ factors through a map $f : C \rightarrow B\text{Symp}(M)$, where C is as above. This map f depends on the choice of a symplectic trivialization of π over the 1-skeleton B_1 of B . We now show that f can be chosen so that (ii) holds for the associated symplectic bundle $Q_f \rightarrow C$.

We need to show that the differentials $(d_C)^{0,2}_2, (d_C)^{0,2}_3$ in the spectral sequence for $Q_f \rightarrow C$ both vanish on $[\omega]$. Let

$$\cdots \rightarrow C_k(B) \xrightarrow{\partial} C_{k-1}(B) \rightarrow \cdots$$

be the *cellular chain complex* for B , and choose 2-cells e_1, \dots, e_k in B whose attaching maps $\alpha_1, \dots, \alpha_k$ form a basis over \mathbf{Q} for the image of ∂

in $C_1(B)$. Then the obvious maps $C_k(B) \rightarrow C_k(C)$ (which are the identity for $k > 1$) give rise to an isomorphism

$$H_2(B, \mathbf{Q}) \bigoplus \bigoplus_i \mathbf{Q}[e_i] \cong H_2(C, \mathbf{Q}).$$

By the naturality of spectral sequences, the vanishing of $(d_B)_2^{0,2}([\omega])$ implies that $(d_C)_2^{0,2}([\omega])$ vanishes on all cycles in $H_2(C, \mathbf{Q})$ coming from $H_2(B, \mathbf{Q})$. Therefore we just need to check that it vanishes on the cycles e_i . For this, we have to choose the trivialization over B_1 so that its pullback by each α_i gives rise to a Hamiltonian bundle over e_i . For this it would suffice that its pullback by each α_i is the “natural trivialization”, i.e the one that extends over the 2-cell e_i . To arrange this, choose any symplectic trivialization over $B_1 = \vee_j \gamma_j$. Then comparing this with the natural trivialization gives rise to a homomorphism

$$\Phi : \bigoplus_i \mathbb{Z}e_i \rightarrow \pi_1 \text{Sym}(M, \mathbb{Z}) \xrightarrow{\text{Flux}} H^1(M, \mathbb{R}).$$

Since the boundary map $\bigoplus_i \mathbb{Z}e_i \rightarrow C_1(B) \otimes \mathbf{Q}$ is injective, we can now change the chosen trivializations over the 1-cells γ_j in B_1 to make $\Phi = 0$.

This ensures that $d_2^{0,2} = 0$ in the bundle over C . Since the map $H^q(C) \rightarrow H^q(B)$ is an isomorphism when $q \geq 3$, the vanishing of $d_3^{0,2}$ for B implies that it vanishes for C . Therefore (ii) holds for $Q \rightarrow C$. By the previous result, this implies that the structural group of $Q \rightarrow C$ reduces to $\text{Ham}(M)$. Therefore, the same holds for $P \rightarrow B$.

So we have established the following result [Lalonde and McDuff (2002)]: Let C be the CW complex obtained by collapsing the 1-skeleton of B to a point and $f : B \rightarrow C$ be the obvious map. Then any Hamiltonian bundle $P \rightarrow B$ is the pullback by f of some Hamiltonian bundle over C .

The above theorem shows that there are two obstructions to the existence of a Hamiltonian structure on a symplectic bundle. Firstly, the bundle must be symplectically trivial over the 1-skeleton B_1 , and secondly the symplectic class on the fiber must extend. The first obstruction obviously depends on the 1-skeleton B_1 while the second, in principle, depends on its 3-skeleton (since we need d_2 and d_3 to vanish on $[\omega]$). However, in fact, it only depends on the 2-skeleton, as is shown in the following lemma: Every symplectic bundle over a 2-connected base B is Hamiltonian. Note that we just have to show that $d_3^{0,2}([\omega]) = 0$. Alternatively, let Sym_0 (resp. $\widetilde{\text{Ham}}$) denote the universal cover of the group $\text{Sym}_0 = \text{Sym}_0(M, \omega)$ (resp

$Ham(M)$), and set $\pi_S = \pi_1(Symp_0)$ so that there are fibrations

$$\widetilde{Ham} \rightarrow Symp_0 \xrightarrow{Flux} H^1(M, \mathbb{R}), \quad B(\pi_S) \rightarrow B Symp_0 \rightarrow B Symp_0.$$

The existence of the first fibration shows that \widetilde{Ham} is homotopy equivalent to $Symp_0$ so that $B\widetilde{Ham} \simeq B Symp_0$, while the second implies that there is a fibration

$$B Symp_0 \rightarrow B Symp_0 \rightarrow K(\pi_S, 2),$$

where $K(\pi_S, 2)$ is an *Eilenberg–MacLane space*. A symplectic bundle over B is equivalent to a homotopy class of maps $B \rightarrow B Symp_0$. If B is 2-connected, the composite $B \rightarrow B Symp_0 \rightarrow K(\pi_S, 2)$ is null homotopic, so that the map $B \rightarrow B Symp_0$ lifts to $B Symp_0$ and hence to the homotopic space $B\widetilde{Ham}$. Composing this map $B \rightarrow B\widetilde{Ham}$ with the projection $B\widetilde{Ham} \rightarrow B Ham$ we get a Hamiltonian structure on the given bundle over B .

Equivalently, use the existence of the fibration $\widetilde{Ham} \rightarrow Symp_0 \rightarrow H^1(M, \mathbb{R})$ to deduce that the subgroup $\pi_1(Ham)$ of \widetilde{Ham} injects into $\pi_1(Symp_0)$. This implies that the relative homotopy groups $\pi_i(Symp_0, Ham)$ vanish for $i > 1$, so that

$$\pi_i(B Symp_0, B Ham) = \pi_{i-1}(Symp_0, Ham) = 0, \quad (i > 2).$$

The desired conclusion now follows by *obstruction theory*. The second proof does not directly use the sequence

$$0 \rightarrow Ham \rightarrow Symp_0 \rightarrow H^1/\Gamma_\omega \rightarrow 0,$$

since the flux group Γ_ω may not be a discrete subgroup of H^1 .

4.12.1.9 Classification of Hamiltonian Structures

The previous subsection discussed the question of the existence of Hamiltonian structures on a given bundle. We now look at the problem of describing and classifying them.

Let $\pi : P \rightarrow B$ be a symplectic bundle satisfying the above conditions and fix an identification of (M, ω) with (M_{b_0}, ω_{b_0}) . Let a be any closed extension of $[\omega]$, $\gamma_1, \dots, \gamma_k$ be a set of generators of the first rational homology group of B , $\{c_i\}$ the dual basis of $H^1(B)$ and T_1, \dots, T_k symplectic trivializations round the γ_i . Assume for the moment that each class $f(T_i, a) \in H^1(M_{b_0}) = H^1(M)$ has an extension $\tilde{f}(T_i, a)$ to P . Subtracting

from a the class $\sum_{i=1}^k \pi^*(c_i) \cup \tilde{f}(T_i, a)$, we get a closed extension a' whose corresponding classes $f(T_i, a')$ belong to Γ_ω .

There remains to prove that the extensions of the $f(T_i, a)$'s exist in Hamiltonian bundles. It is enough to prove that the fiber inclusion $M \rightarrow P$ induces an injection on the first homology group. One only needs to prove this over the 2-skeleton B_2 of B , and we can assume as well that B_2 is a wedge of 2-spheres. Hence this is a consequence of the easy fact that the evaluation of a Hamiltonian loop on a point of M gives a 1-cycle of M that is trivial in rational homology, i.e., that the differential $d_2^{0,1}$ vanishes in the cohomology spectral sequence for $P \rightarrow B$ (see [Lalonde *et al.* (1999)], where this is proved by elementary methods).

The following result extends the above lemma: Let $P \rightarrow B$ be a symplectic bundle with a given symplectic trivialization of P over B_1 , and let $a \in H^2(P)$ be a normalized extension of the fiber symplectic class. Then the restriction of a to $\pi^{-1}(B_1)$ defines and is defined by a homomorphism Φ from $\pi_1(B)$ to Γ_ω .

As above, we can use the given trivialization to identify the holonomy round some loop $s \mapsto \gamma(s) \in B_1$ with a family of symplectomorphisms $\Phi_s^\gamma : M \rightarrow M, s \in [0, 1]$. Given a 1-cycle $\delta : S^1 \rightarrow M$ in the fiber M over $1 \in \partial D^2$, consider the closed 2-cycle $C(\gamma, \delta) = c_1 \cup c_2$ as before. Since $\tau(c_1) = 0$,

$$\tau(C(\gamma, \delta)) = \tau(c_2) = -\text{Flux}(\{\Phi_s^\gamma\})(\delta).$$

If we now set

$$\Phi(\gamma) = -\text{Flux}(\{\Phi_s^\gamma\}),$$

it is easy to check that Φ is a homomorphism [Lalonde and McDuff (2002)]. Its values are in Γ_ω by the definition of normalized extension classes.

The next task is to prove the above theorem that characterizes Hamiltonian structures. Thus we need to understand the homotopy classes of lifts \tilde{g} of the classifying map $g : B \rightarrow B \text{Symp}(M, \omega)$ of the underlying symplectic bundle to $B \text{Ham}(M)$. We first consider the intermediate lift \hat{g} of g into $B \text{Symp}_0(M, \omega)$. In view of the fibration sequence

$$\pi_0(\text{Symp}) \rightarrow B \text{Symp}_0 \rightarrow B \text{Symp} \rightarrow B(\pi_0(\text{Symp}))$$

in which each space is mapped to the homotopy fiber of the subsequent map, a map $g : B \rightarrow B \text{Symp}$ lifts to $\hat{g} : B \rightarrow B \text{Symp}_0$ iff the symplectic bundle given by g can be trivialized over the 1-skeleton B_1 of B . Moreover such

lifts are in bijective correspondence with the elements of $\pi_0(\text{Symp})$ and so correspond to an identification (up to symplectic isotopy) of (M, ω) with the fiber (M_{b_0}, ω_{b_0}) at the base point b_0 (recall that B is always assumed to be connected).

To understand the full lift \tilde{g} , recall the above exact sequence

$$\{\text{Id}\} \longrightarrow \text{Ham}(M, \omega) \longrightarrow \text{Symp}_0(M, \omega) \xrightarrow{\text{Flux}} H^1(M, \mathbb{R})/\Gamma_\omega \longrightarrow \{0\}.$$

If Γ_ω is discrete, then the space $H^1(M, \mathbb{R})/\Gamma_\omega$ is homotopy equivalent to a torus and we can investigate the liftings \tilde{g} by homotopy theoretic arguments about the fibration

$$H^1(M, \mathbb{R})/\Gamma_\omega \rightarrow B\text{Ham}(M, \omega) \rightarrow B\text{Symp}_0(M, \omega).$$

Now, suppose that a symplectic bundle $\pi : P \rightarrow B$ is given that satisfies the above conditions. Fix an identification of (M, ω) with (M_{b_0}, ω_{b_0}) . We have to show that lifts from $B\text{Symp}_0$ to $B\text{Ham}$ are in bijective correspondence with equivalence classes of normalized extensions a of the fiber symplectic class. There is a lift iff there is a normalized extension class a . Therefore, it remains to show that the equivalence relations correspond. The essential reason why this is true is that the induced map

$$\pi_i(\text{Ham}(M, \omega)) \rightarrow \pi_i(\text{Symp}_0(M, \omega))$$

is an injection for $i = 1$ and an isomorphism for $i > 1$. This, in turn, follows from the exactness of the sequence (*).

$$\{\text{Id}\} \longrightarrow \text{Ham}(M, \omega) \longrightarrow \text{Symp}_0(M, \omega) \xrightarrow{\text{Flux}} H^1(M, \mathbb{R})/\Gamma_\omega \longrightarrow \{0\}.$$

Let us spell out a few more details, first when B is simply connected. Then the classifying map from the 2-skeleton B_2 to $B\text{Symp}_0$ has a lift to $B\text{Ham}$ iff the image of the induced map

$$\pi_2(B_2) \rightarrow \pi_2(B\text{Symp}_0(M)) = \pi_1\text{Symp}_0(M)$$

lies in the kernel of the *flux homomorphism*

$$\text{Flux} : \pi_1(\text{Symp}_0(M)) \longrightarrow \Gamma_\omega.$$

Since $\pi_1(\text{Ham}(M, \omega))$ injects into $\pi_1(\text{Symp}_0(M, \omega))$, there is only one such lift up to homotopy. Standard arguments now show that this lift can be extended uniquely to the rest of B . Hence in this case there is a unique lift. Correspondingly there is a unique equivalence class of extensions a [Lalonde and McDuff (2002)].

Now let us consider the general case: We are given a map $g : B \rightarrow B \operatorname{Symp}_0$ and want to identify the different homotopy classes of liftings of g to $B \operatorname{Ham}$. Let $C = B/B_1$ as above. There is a symplectic trivialization T over B_1 that is compatible with the given identification of the base fiber and induces a map $C \rightarrow B \operatorname{Symp}_0$ which lifts to $B \operatorname{Ham}$. Since this lifting $g_{T,C}$ of $C \rightarrow B \operatorname{Symp}_0$ is unique, each isotopy class T of such trivializations over B_1 gives rise to a unique homotopy class g_T of maps $B \rightarrow B \operatorname{Ham}$, namely

$$g_T : B \longrightarrow C \xrightarrow{g_{T,C}} B \operatorname{Ham}.$$

Note that g_T is a lifting of f and that every lifting occurs this way.

Standard arguments show that two such isotopy classes differ by a homomorphism

$$\pi_1(B) \longrightarrow \pi_1(\operatorname{Symp}_0).$$

Moreover, the corresponding maps g_T and g'_T are homotopic iff T and T' differ by a homomorphism with values in $\pi_1(\operatorname{Ham})$. Thus homotopy classes of liftings of g to $B \operatorname{Ham}$ are classified by homomorphisms $\pi_1(B) \rightarrow \Gamma_\omega$. These homomorphisms are precisely what defines the equivalence classes of extensions a .

4.12.2 Properties of General Hamiltonian Bundles

The key to extending results about Hamiltonian bundles over S^2 to higher-dimensional bases is their functorial properties, in particular their behavior under composition. Before discussing this, it is useful to establish the fact that this class of bundles is stable under small perturbations of the symplectic form on M [Lalonde and McDuff (2002)].

4.12.2.1 Stability

Moser's argument implies that *for every symplectic structure ω on M there is a Serre fibration*

$$\operatorname{Symp}(M, \omega) \longrightarrow \operatorname{Diff}(M) \longrightarrow \mathcal{S}_\omega,$$

where \mathcal{S}_ω is the space of all symplectic structures on M that are diffeomorphic to ω . At the level of classifying spaces, this gives a *homotopy fibration*

$$\mathcal{S}_\omega \hookrightarrow B \operatorname{Symp}(M, \omega) \longrightarrow B \operatorname{Diff}(M).$$

Any smooth fiber bundle $P \rightarrow B$ with fiber M is classified by a map $B \rightarrow B\operatorname{Diff}(M)$, and isomorphism classes of symplectic structures on it with fiber (M, ω) correspond to homotopy classes of sections of the associated fibration $W(\omega) \rightarrow B$ with fiber \mathcal{S}_ω . We will suppose that π is described by a finite set of local trivializations $T_i : \pi^{-1}(V_i) \rightarrow V_i \times M$ with the transition functions $\phi_{ij} : V_i \cap V_j \rightarrow \operatorname{Diff}(M)$.

Suppose that $M \rightarrow P \rightarrow B$ is a smooth fibration constructed from a cocycle (T, ϕ_{ij}) with the following property: there is a symplectic form ω on M such that for each $x \in M$ the convex hull of the finite set of forms $\{\phi_{ij}^*(\omega) : x \in V_i \cap V_j\}$ lies in the set \mathcal{S}_ω of symplectic forms diffeomorphic to ω . Then $(M, \omega) \rightarrow P \rightarrow B$ may be given the structure of a symplectic bundle [Lalonde and McDuff (2002)].

It suffices to construct a section σ of $W(\omega) \rightarrow B$. The hypothesis implies that for each x the convex hull of the set of forms $T_i(x)^*(\omega), x \in V_i$, lies in the fiber of $W(\omega)$ at x . Hence we may take $\sigma(x) = \rho_i T_i^*(\omega)$, where ρ_i is a partition of unity subordinate to the cover V_i .

Let $P \rightarrow B$ be a symplectic bundle with closed fiber (M, ω) and compact base B . There is an open neighborhood of ω in the space (M) of all symplectic forms on M such that, for all $\omega' \in U$, the structural group of $\pi : P \rightarrow B$ may be reduced to $\operatorname{Symp}(M, \omega')$.

Trivialize $P \rightarrow B$ so that $\phi_{ij}^*(\omega) = \omega$ for all i, j . Then the hypothesis of the lemma is satisfied for all ω' sufficiently close to ω by the openness of the symplectic condition.

Thus the set $\pi(M)$ of symplectic forms on M , with respect to which π is symplectic, is open. The aim of this section is to show that a corresponding statement is true for Hamiltonian bundles. The following example shows that the Hamiltonian property need not survive under *large* perturbations of ω .

Here is an example of a smooth family of symplectic bundles that is Hamiltonian at all times $0 \leq t < 1$ but is non-Hamiltonian at time 1. Let h_t ($0 \leq t \leq 1$) be a family of diffeomorphisms of M with $h_0 = \operatorname{Id}$ and define

$$Q = M \times [0, 1] \times [0, 1] / (x, 1, t) \equiv (h_t(x), 0, t).$$

Thus we can think of Q as a family of bundles $\pi : P_t \rightarrow S^1$ with monodromy h_t at time t . It was shown in [Seidel (1997)] that there are smooth families of symplectic forms ω_t and diffeomorphisms $h_t \in \operatorname{Symp}(M, \omega_t)$ for $t \in [0, 1]$ such that h_t is not in the identity component of $\operatorname{Symp}(M, \omega_t)$ for $t = 1$ but is in this component for $t < 1$. For such h_t each bundle $P_t \rightarrow S^1$ is symplectic. Moreover, it is symplectically trivial and hence Hamiltonian

for $t < 1$ but is non-Hamiltonian at $t = 1$.

We have the following result [Lalonde and McDuff (2002)]: A Hamiltonian bundle $\pi : P \rightarrow B$ is stable iff the restriction map $H^2(P) \rightarrow H^2(M)$ is surjective. If $\pi : P \rightarrow B$ is Hamiltonian with respect to ω' then $[\omega']$ is in the image of $H^2(P) \rightarrow H^2(M)$. If π is stable, then $[\omega']$ fills out a neighborhood of $[\omega]$ which implies surjectivity. Conversely, suppose that we have surjectivity. To check (i) let $\gamma : S^1 \rightarrow B$ be a loop in B and suppose that $\gamma^*(P)$ is identified symplectically with the product bundle $S^1 \times (M, \omega)$. Let $\omega_t, 0 \leq t \leq \varepsilon$, be a (short) smooth path with $\omega_0 = \omega$. Then, because $P \rightarrow B$ has the structure of an ω_t -symplectic bundle for each t , each fiber M_b has a corresponding smooth family of symplectic forms $\omega_{b,t}$ of the form $g_{b,t}^* \psi_b^*(\omega_t)$, where ψ_b is a symplectomorphism $(M_b, \omega_b) \rightarrow (M, \omega)$. Hence, for each t , $\gamma^*(P)$ can be symplectically identified with

$$\cup_{s \in [0,1]} (\{s\} \times (M, g_{s,t}^*(\omega_t))), \quad \text{where} \quad g_{1,t}^*(\omega_t) = \omega_t$$

and the $g_{s,t}$ lie in an arbitrarily small neighborhood U of the identity in $Diff(M)$. By Moser's homotopy argument, we can choose U so small that each $g_{1,t}$ is isotopic to the identity in the group $Symp(M, \omega_t)$.

Now, the pullback of a stable Hamiltonian bundle is stable [Lalonde and McDuff (2002)]. Suppose that $P \rightarrow B$ is the pullback of $P' \rightarrow B'$ via $B \rightarrow B'$ so that there is a commutative diagram

$$\begin{array}{ccc} P & \xrightarrow{\quad} & P' \\ \downarrow & & \downarrow \\ B & \xrightarrow{\quad} & B' \end{array}$$

By hypothesis, the restriction $H^2(P') \rightarrow H^2(M)$ is surjective. But this map factors as $H^2(P') \rightarrow H^2(P) \rightarrow H^2(M)$. Hence $H^2(P) \rightarrow H^2(M)$ is also surjective. We have the following lemma:

- (i) Every Hamiltonian bundle over S^2 is stable.
- (ii) Every symplectic bundle over a 2-connected base B is Hamiltonian stable.

(i) holds because every Hamiltonian bundle over S^2 is c -split, in particular the restriction map $H^2(P) \rightarrow H^2(M)$ is surjective.

The above theorem states that *every Hamiltonian bundle is stable*. To prove this, first observe that we can restrict to the case when B is simply connected. For the map $B \rightarrow BHam(M)$ classifying P factors through a map $C \rightarrow BHam(M)$, where $C = B/B_1$ as before, and the stability of the

induced bundle over C implies that for the original bundle.

Next observe that a Hamiltonian bundle $P \rightarrow B$ is stable iff the differentials $d_k^{0,2} : E_k^{0,2} \rightarrow E_k^{k,3-k}$ in its Leray cohomology spectral sequence vanish on the whole of $H^2(M)$ for $k = 2, 3$. We can reduce the statement for $d_2^{0,2}$ to the case $B = S^2$. Thus $d_2^{0,2} = 0$. Similarly, we can reduce the statement for $d_3^{0,2}$ to the case $B = S^3$.

4.12.2.2 Functorial Properties

We begin with some trivial observations and then discuss composites of Hamiltonian bundles. The first lemma is true for any class of bundles with specified structural group [Lalonde and McDuff (2002)].

Suppose that $\pi : P \rightarrow B$ is Hamiltonian and that $g : B' \rightarrow B$ is a continuous map. Then the induced bundle $\pi' : g^*(P) \rightarrow B'$ is Hamiltonian.

Recall that any extension τ of the forms on the fibers is called a connection form.

If $P \rightarrow B$ is a smooth Hamiltonian fiber bundle over a symplectic base (B, σ) and if P is compact then there is a connection form Ω^κ on P that is symplectic.

The bundle P carries a closed connection form τ . Since P is compact, the form $\Omega^\kappa = \tau + \kappa\pi^*(\sigma)$ is symplectic for large κ .

Observe that the deformation type of the form Ω^κ is unique for sufficiently large κ since given any two closed connection forms τ, τ' the *linear isotopy*

$$t\tau + (1-t)\tau' + \kappa\pi^*(\sigma), \quad 0 \leq t \leq 1,$$

consists of symplectic forms for sufficiently large κ . However, it can happen that there is a symplectic connection form τ such that $\tau + \kappa\pi^*(\sigma)$ is not symplectic for small $\kappa > 0$, even though it is symplectic for large κ . (For example, suppose $P = M \times B$ and that τ is the sum $\omega + \pi^*(\omega_B)$ where $\omega_B + \sigma$ is not symplectic.)

Let us now consider the behavior of Hamiltonian bundles under composition. If

$$(M, \omega) \rightarrow P \xrightarrow{\pi_P} X, \quad \text{and} \quad (F, \sigma) \rightarrow X \xrightarrow{\pi_X} B$$

are Hamiltonian fiber bundles, then the restriction

$$\pi_P : W = \pi_P^{-1}(F) \longrightarrow F$$

is a Hamiltonian fiber bundle. Since F is a manifold, we can assume without loss of generality that $W \rightarrow F$ is smooth. Moreover, since (F, σ) is symplectic, the manifold W carries a symplectic connection form Ω_W^κ , and it is natural to ask when the composite map $\pi : P \rightarrow B$ with fiber (W, Ω_W^κ) is itself Hamiltonian.

Suppose in the above situation that B is simply connected and that P is compact. Then $\pi = \pi_X \circ \pi_P : P \rightarrow B$ is a Hamiltonian fiber bundle with fiber (W, Ω_W^κ) , where

$$\Omega_W^\kappa = \tau_W + \kappa \pi_P^*(\sigma),$$

τ_W is any *symplectic connection form* on W , and κ is sufficiently large.

By above lemma, we may assume that the base B as well as the fibrations are smooth. We first show that there is some symplectic form on W for which π is Hamiltonian and then show that it is Hamiltonian with respect to the given form Ω_W^κ .

Let τ_P (resp. τ_X) be a closed connection form with respect to the bundle π_P , (resp. π_X), and let τ_W be its restriction to W . Then Ω_W^κ is the restriction to W of the closed form

$$\Omega_P^\kappa = \tau_P + \kappa \pi_P^*(\tau_X).$$

By increasing κ if necessary we can ensure that Ω_P^κ restricts to a symplectic form on every fiber of π not just on the chosen fiber W . This shows firstly that $\pi : P \rightarrow B$ is symplectic, because there is a well defined symplectic form on each of its fibers, and secondly that it is Hamiltonian with respect to this form Ω_W^κ on the fiber W . Hence $H^2(P)$ surjects onto $H^2(W)$.

Now suppose that τ_W is *any* closed connection form on $\pi_P : W \rightarrow F$. Because the restriction map $H^2(P) \rightarrow H^2(W)$ is surjective, the cohomology class $[\tau_W]$ is the restriction of a class on P and so, by Thurston's construction, the form τ_W can be extended to a closed connection form τ_P for the bundle π_P . Therefore the previous argument applies in this case too.

We know that the restriction map $H^2(P) \rightarrow H^2(M)$ is surjective, so that the cohomology class $[\tau_W]$ of the connection form τ_W of $W \rightarrow F$ is the restriction of a class on P , and by Thurston's construction the form τ_W can be extended to a closed connection form τ_P for the bundle π_P . Choose also a closed connection form τ_X for the bundle π_X . Then Ω_W^κ is the restriction to W of the closed form $\Omega_P^\kappa = \tau_P + \kappa \pi_P^*(\tau_X)$. By increasing κ if necessary we can ensure that Ω_P^κ restricts to a symplectic form on

every fiber of π not just on the the chosen fiber W . This shows firstly that $\pi : P \rightarrow B$ is symplectic, because there is a well defined symplectic form on each of its fibers, and secondly that it is Hamiltonian with respect to this form Ω_W^κ on the fiber W . Hence $H^2(P)$ surjects onto $H^2(W)$.

Now suppose that τ_W is *any* closed connection form on $\pi_P : W \rightarrow F$. Because the restriction map $H^2(P) \rightarrow H^2(W)$ is surjective, the cohomology class $[\tau_W]$ is the restriction of a class on P and so, by Thurston's construction, the form τ_W can be extended to a closed connection form τ_P for the bundle π_P . Therefore the previous argument applies in this case too.

Now let us consider the general situation, when $\pi_1(B) \neq 0$. The proof of the lemma above applies to show that the composite bundle $\pi : P \rightarrow B$ is symplectic with respect to suitable Ω_W^κ and that it has a symplectic connection form. However, even though $\pi_X : X \rightarrow B$ is symplectically trivial over the 1-skeleton of B the same may not be true of the composite map $\pi : P \rightarrow B$. Moreover, in general it is not clear whether triviality with respect to one form Ω_W^κ implies it for another. Therefore, we may conclude the following: If

$$(M, \omega) \rightarrow P \xrightarrow{\pi_P} X \quad \text{and} \quad (F, \sigma) \rightarrow X \xrightarrow{\pi_X} B$$

are Hamiltonian fiber bundles and P is compact, then the composite $\pi = \pi_X \circ \pi_P : P \rightarrow B$ is a symplectic fiber bundle with respect to any form Ω_W^κ on its fiber $W = \pi^{-1}(pt)$, provided that κ is sufficiently large. Moreover if π is symplectically trivial over the 1-skeleton of B with respect to Ω_W^κ then π is Hamiltonian.

In practice, we will apply these results in cases where $\pi_1(B) = 0$. We will not specify the precise form on W , assuming that it is Ω_W^κ for a suitable κ .

4.12.2.3 Splitting of Rational Cohomology

We write $H_*(X), H^*(X)$ for the rational (co)homology of X . Recall that a bundle $\pi : P \rightarrow B$ with fiber M is said to be *c-split* if

$$H^*(P) \cong H^*(B) \otimes H^*(M).$$

This happens iff $H_*(M)$ injects into $H_*(P)$. Dually, it happens iff the restriction map $H^*(P) \rightarrow H^*(M)$ is onto. Note also that a bundle $P \rightarrow B$ *c-splits* iff the E_2 term of its cohomology spectral sequence is a product and all the differentials $d_k, k \geq 2$, vanish [Lalonde and McDuff (2002)].

The following *surjection lemma* is obvious but useful: Consider a commutative diagram

$$\begin{array}{ccc} P' & \xrightarrow{\quad} & P \\ \downarrow & & \downarrow \\ B' & \xrightarrow{\quad} & B \end{array}$$

where P' is the induced bundle. Then:

- (i) If $P \rightarrow B$ is c -split so is $P' \rightarrow B'$.
- (ii) If $P' \rightarrow B'$ is c -split and $H_*(B') \rightarrow H_*(B)$ is surjective, then $P \rightarrow B$ is c -split.

To prove (i), we can use the fact that $P \rightarrow B$ is c -split iff the map $H_*(M) \rightarrow H_*(P)$ is injective. To prove (ii), the induced map on the E_2 -term of the cohomology spectral sequences is injective. Therefore the existence of a nonzero differential in the spectral sequence $P \rightarrow B$ implies one for the pullback bundle $P' \rightarrow B'$.

As a corollary, suppose that $P \rightarrow W$ is a Hamiltonian fiber bundle over a symplectic manifold W and that its pullback to some blowup \widehat{W} of W is c -split. Then $P \rightarrow W$ is c -split.

This follows immediately from (ii) above since the map $H_*(\widehat{W}) \rightarrow H_*(W)$ is surjective.

If $(M, \omega) \xrightarrow{\pi} P \rightarrow B$ is a compact Hamiltonian bundle over a simply connected CW-complex B and if every Hamiltonian fiber bundle over M and B is c -split, then every Hamiltonian bundle over P is c -split.

Let $\pi_E : E \rightarrow P$ be a Hamiltonian bundle with fiber F and let $F \rightarrow W \rightarrow M$ be its restriction over M . Then by assumption the latter bundle c -splits so that $H_*(F)$ injects into $H_*(W)$. The above composition lemma implies that the composite bundle $E \rightarrow B$ is Hamiltonian with fiber W and therefore also c -splits. Hence $H_*(W)$ injects into $H_*(E)$. Thus $H_*(F)$ injects into $H_*(E)$, as required.

If Σ is a closed orientable surface then any Hamiltonian bundle over $S^2 \times \dots \times S^2 \times \Sigma$ is c -split.

Consider any degree one map f from $\Sigma \rightarrow S^2$. Because $Ham(M, \omega)$ is connected, $BHam(M, \omega)$ is simply connected, and therefore any homotopy class of maps from $\Sigma \rightarrow BHam(M, \omega)$ factors through f . Thus any Hamiltonian bundle over Σ is the pullback by f of a Hamiltonian bundle over S^2 . Because such bundles c -split over S^2 , the same is true over Σ .

Any Hamiltonian bundle over $S^2 \times \dots \times S^2 \times S^1$ is c -split. For each

$k \geq 1$, every Hamiltonian bundle over S^k c -splits.

There is for each k a k -dimensional closed manifold X such that every Hamiltonian bundle over X c -splits. Given any Hamiltonian bundle $P \rightarrow S^k$ consider its pullback to X by a map $f : X \rightarrow S^k$ of degree 1. Since the pullback c -splits, the original bundle does too by the surjection lemma.

As we shall see this result implies that the action of the homology groups of $Ham(M)$ on $H_*(M)$ is always trivial. Here are some other examples of situations in which Hamiltonian bundles are c -split.

Every Hamiltonian bundle over $\mathbf{CP}^{n_1} \times \dots \times \mathbf{CP}^{n_k}$ c -splits.

Let us prove first that it splits over \mathbf{CP}^n . Use induction over n . Again it holds when $n = 1$. Assuming the result for n let us prove it for $n + 1$. Let B be the blowup of \mathbf{CP}^{n+1} at one point. Then B fibers over \mathbf{CP}^n with fiber \mathbf{CP}^1 . Thus every Hamiltonian bundle over B c -splits.

Every Hamiltonian bundle whose structural group reduces to a subtorus $T \subset Ham(M)$ c -splits [Lalonde and McDuff (2002)]. It suffices to consider the universal model

$$M \longrightarrow ET \times_T M \longrightarrow BT,$$

and hence to show that all Hamiltonian bundles over BT are c -split. But this is equal to $\mathbf{CP}^\infty \times \dots \times \mathbf{CP}^\infty$ and the proof that the i^{th} group of homology of the fiber injects in $P \rightarrow \mathbf{CP}^\infty \times \dots \times \mathbf{CP}^\infty$ may be reduced to the proof that it injects in the restriction of the bundle P over $\mathbf{CP}^j \times \dots \times \mathbf{CP}^j$ for a sufficiently large j .

Note that the proof of the above corollary shows that every Hamiltonian bundle over $\mathbf{CP}^\infty \times \dots \times \mathbf{CP}^\infty$ c -splits. Since the structural group of such a bundle can be larger than the torus T , the result presented here extends the Atiyah-Bott splitting theorem for Hamiltonian bundles with structural group T .

For completeness, we show how the above corollary leads to a proof of the splitting of G -equivariant cohomology where G is a Lie subgroup of $Ham(M, \omega)$.

If G is a compact connected Lie group that acts in a Hamiltonian way on M then any bundle $P \rightarrow B$ with fiber M and structural group G is c -split. In particular,

$$H_G^*(M) \cong H^*(M) \otimes H^*(BG).$$

We only need to prove the second statement, since

$$M_G = EG \times_G M \longrightarrow BG$$

is the *universal bundle*. Every compact connected Lie group G is the image of a homomorphism $T \times H \rightarrow G$, where the torus T maps onto the identity component of the center of G and H is the semi-simple Lie group corresponding to the *commutator subalgebra* $[Lie(G), Lie(G)]$ in the Lie algebra $Lie(G)$. It is easy to see that this homomorphism induces a surjection on rational homology $BT \times BH \rightarrow BG$. Therefore, we may suppose that $G = T \times H$. Let $T_{max} = (S^1)^k$ be the maximal torus of the semi-simple group H . Then the induced map on cohomology

$$H^*(BH) \rightarrow H^*(BT_{max}) = \mathbf{Q}[a_1, \dots, a_k]$$

takes $H^*(BH)$ bijectively onto the set of polynomials in $H^*(BT_{max})$ that are invariant under the action of the Weyl group, by the *Borel–Hirzebruch theorem*. Hence the maps

$$BT_{max} \rightarrow BH \quad \text{and} \quad BT \times BT_{max} \rightarrow BG$$

induce a surjection on homology. Therefore the desired result follows from the surjection lemma and the *Atiyah–Bott theorem*.

We have the following lemma [Lalonde and McDuff (2002)]: Every Hamiltonian bundle over a coadjoint orbit c -splits.

This is an immediate consequence of the results by [Grossberg and Karshon (1994)] on Bott towers. Recall that a *Bott tower* is an *iterated fibration of Kähler manifolds*

$$M_k \rightarrow M_{k-1} \rightarrow \cdots \rightarrow M_1 = S^2$$

where each map $M_{i+1} \rightarrow M_i$ is a fibration with fiber S^2 . They show that any coadjoint orbit X can be blown up to a manifold that is diffeomorphic to a Bott tower M_k . Moreover the *blow-down map* $M_k \rightarrow X$ induces a surjection on rational homology. Every Hamiltonian bundle over M_k c -splits. Hence the result follows from the surjection lemma.

Every Hamiltonian bundle over a 3-complex X c -splits. As in the proof of stability given above, we can reduce this to the cases $X = S^2$ and $X = S^3$. The only difference from the stability result is that we now require the differentials $d_2^{0,q}, d_3^{0,q}$ to vanish for all q rather than just at $q = 2$.

Every Hamiltonian bundle over a product of spheres c -splits, provided that there are no more than 3 copies of S^1 .

By hypothesis,

$$B = \prod_{i \in I} S^{2m_i} \times \prod_{j \in J} S^{2n_j+1} \times T^k,$$

where $n_i > 0$ and $0 \leq k \leq 3$. Set

$$B' = \prod_{i \in I} \mathbf{CP}^{m_i} \times \prod_{j \in J} \mathbf{CP}^{n_j} \times T^{|J|} \times T^\ell,$$

where $\ell = k$ if $k + |J|$ is even, and $= k + 1$ otherwise. Since $\mathbf{CP}^{n_i} \times S^1$ maps onto S^{2n_i+1} by a map of degree 1, there is a homology surjection $B' \rightarrow B$ that maps the factor T^ℓ to T^k . By the surjection lemma, it suffices to show that the pullback bundle $P' \rightarrow B'$ is c -split.

Consider the fibration

$$T^{|J|} \times T^\ell \rightarrow B' \rightarrow \prod_{i \in I} \mathbf{CP}^{m_i} \times \prod_{j \in J} \mathbf{CP}^{n_j}.$$

Since $|J| + \ell$ is even, we can think of this as a Hamiltonian bundle. Moreover, by construction, the restriction of the bundle $P' \rightarrow B'$ to $T^{|J|} \times T^\ell$ is the pullback of a bundle over T^k , since the map $T^{|J|} \rightarrow B$ is nullhomotopic. (Note that each S^1 factor in $T^{|J|}$ goes into a different sphere in B .) Because $k \leq 3$, the bundle over T^k c -splits. Hence we can conclude that $P' \rightarrow B'$ c -splits.

Every Hamiltonian bundle whose fiber has cohomology generated by H^2 is c -split. This is an immediate consequence of the stability theorem.

Any Hamiltonian fibration c -splits if its base B is the image of a homology surjection from a product of spheres and projective spaces, provided that there are no more than three S^1 factors. One can also consider iterated fibrations of projective spaces, rather than simply products. However, we have not yet managed to deal with arbitrary products of spheres. In order to do this, it would suffice to show that every Hamiltonian bundle over a torus T^m c -splits. This question has not yet been resolved for $m \geq 4$.

4.12.2.4 Hamiltonian Bundles and Gromov–Witten Invariants

We begin by sketching an alternate proof that every Hamiltonian bundle over $B = \mathbf{CP}^n$ is c -split that generalizes the arguments in [McDuff (2000)]. We will use the language of [McDuff (1999)], which is based on the Liu–Tian [Liu and Tian (1998)] approach to general *Gromov–Witten invariants*. Clearly, any treatment of general Gromov–Witten invariants could be used instead.

We have the fundamental result: [Lalonde and McDuff (2002)] *Every Hamiltonian bundle over CP^n is c -split*. To prove this, the basic idea is to show that the inclusion $\iota : H_*(M) \rightarrow H_*(P)$ is injective by showing that for

every nonzero $a \in H_*(M)$ there is $b \in H_*(M)$ and $\sigma \in H_2(P; \mathbb{Z})$ for which the Gromov–Witten invariant $n_P(\iota(a), \iota(b); \sigma)$ is nonzero. Intuitively speaking, this invariant ‘counts the number of isolated J –holomorphic curves in P that represent the class σ and meet the classes $\iota(a), \iota(b)$.’ More correctly, it is defined to be the intersection number of the image of the evaluation map

$$\text{ev} : \overline{\mathcal{M}}_{0,2}^\nu(P, J, \sigma) \longrightarrow P \times P,$$

with the class $\iota(a) \times \iota(b)$, where $\overline{\mathcal{M}}_{0,2}^\nu(P, J, \sigma)$ is a virtual moduli cycle made from perturbed J –holomorphic curves with 2 marked points, and ev is given by evaluating at these two points. As explained in [McDuff (1999); McDuff (2000)], $\overline{\mathcal{M}}^\nu = \overline{\mathcal{M}}_{0,2}^\nu(P, J, \sigma)$ is a branched pseudo-manifold, i.e., a kind of stratified space whose top dimensional strata are oriented and have rational labels. Roughly speaking, one can think of it as a finite simplicial complex, whose dimension d equals the ‘*formal dimension*’ of the moduli space, i.e., the *index* of the relevant operator. The elements of $\overline{\mathcal{M}}^\nu$ are stable maps $[\Sigma, h, z_1, z_2]$ where z_1, z_2 are two marked points on the nodal, genus 0, Riemann surface Σ , and the map $h : \Sigma \rightarrow P$ satisfies a *perturbed Cauchy–Riemann equation* $\overline{\mathcal{M}}_J h = \nu_h$. The perturbation ν can be arbitrarily small, and is chosen so that each stable map in $\overline{\mathcal{M}}^\nu$ is a regular point for the appropriate Fredholm operator. Hence $\overline{\mathcal{M}}^\nu$ is often called a *regularization* of the unperturbed moduli space $\overline{\mathcal{M}} = \overline{\mathcal{M}}_{0,2}(P, J, \sigma)$ of all J –holomorphic stable maps.

Given any Hamiltonian bundle $P_S \rightarrow S^2$ and any $a \in H_*(M)$, it was shown in [Lalonde *et al.* (1999); McDuff (2000)] that there is $b \in H_*(M)$ and a lift $\sigma_S \in H_2(P_S; \mathbb{Z})$ of the fundamental class of S^2 to P_S such that

$$n_{P_S}(\iota_S(a), \iota_S(b); \sigma_S) \neq 0,$$

where ι_S denotes the inclusion into P_S . Therefore, if P_S is identified with the restriction of P to a complex line L_0 in B and if a, b and σ_S are as above, it suffices to prove that

$$n_{P_S}(\iota_S(a), \iota_S(b); \sigma_S) = n_P(\iota(a), \iota(b); \sigma)$$

where σ is the image of σ_S in P . Note that a direct count shows that the dimensions of the appropriate virtual moduli spaces $\overline{\mathcal{M}}_{0,2}^\nu(P_S, J_S, \sigma_S)$ and $\overline{\mathcal{M}}_{0,2}^\nu(P, J, \sigma)$ differ by the codimension of $P_S \times P_S$ in $P \times P$ (which equals the codimension of $\mathbf{CP}^1 \times \mathbf{CP}^1$ in $\mathbf{CP}^n \times \mathbf{CP}^n$) so that the both sides are well-defined.

As was shown in [McDuff (2000)], one can construct the virtual moduli cycle $\overline{\mathcal{M}}^\nu(P_S) = \overline{\mathcal{M}}_{0,2}^\nu(P_S, J_S, \sigma_S)$ using an *almost complex structure* J_S and a *perturbation* ν that are compatible with the bundle. In particular, this implies that the projection $P_S \rightarrow S^2$ is (J_S, j) -holomorphic (where j is the usual complex structure on S^2) and that every element of $\overline{\mathcal{M}}^\nu(P_S)$ projects to a j -holomorphic stable map in S^2 .

Following [Lalonde and McDuff (2002)], we claim that this is also true for the bundle $P \rightarrow B$. In other words, we can choose J so that the projection $(P, J) \rightarrow (B, j)$ is holomorphic, where j is the usual complex structure on $B = \mathbf{CP}^n$, and we can choose ν so that every element in $\overline{\mathcal{M}}^\nu(P)$ projects to a j -holomorphic stable map in B . The proof is exactly as before [McDuff (2000)]. The essential point is that every element of the unperturbed moduli space $\overline{\mathcal{M}}_{0,2}(\mathbf{CP}^n, j, [\mathbf{CP}^1])$ is regular. In fact, the top stratum in $\overline{\mathcal{M}}_{0,2}(\mathbf{CP}^n, j, [\mathbf{CP}^1])$ is the space $\mathcal{L} = \overline{\mathcal{M}}_{0,2}(\mathbf{CP}^n, j, [\mathbf{CP}^1])$ of all lines in \mathbf{CP}^n with 2 distinct marked points. The other stratum completes this space by adding in the lines with two coincident marked points, which are represented as stable maps by a line together with a ghost bubble containing the two points.

It follows that there is a projection map

$$proj: \overline{\mathcal{M}}_{0,2}^\nu(P, J, \sigma) \longrightarrow \overline{\mathcal{M}}_{0,2}(\mathbf{CP}^n, j, [\mathbf{CP}^1]).$$

Moreover the inverse image of a line $L \in$ can morally speaking be identified with $\overline{\mathcal{M}}_{0,2}^\nu(P_S, J_S, \sigma_S)$. The latter statement would be correct if we were considering ordinary moduli spaces of stable maps, but the virtual moduli space is not usually built in such a way that the fibers $(proj)^{-1}(L)$ have the needed structure of a branched pseudo-manifold. However, we can choose to construct $\overline{\mathcal{M}}_{0,2}^\nu(P, J, \sigma)$ so that this is true for all lines near a fixed line L_0 . A detailed recipe is given in [McDuff (2000)] for constructing $\overline{\mathcal{M}}^\nu$ from the unperturbed moduli space $\overline{\mathcal{M}}$. The construction is based on the choice of suitable covers $\{U_i\}, \{V_I\}$ of $\overline{\mathcal{M}}$ and of perturbations ν_i over each U_i . Because regularity is an open condition, one can make these choices first for all stable maps that project to the fixed line L_0 and then extend to the set of stable maps that project to nearby lines in such a way that $\overline{\mathcal{M}}^\nu$ is locally a product near the fiber over L_0 (see the proof of [McDuff (2000)] for a very similar construction).

Once this is done, the rest of the argument is easy. If we identify P_S with $\pi^{-1}(L_0)$ and choose a representative $\alpha \times \beta$ of $\iota_S(a) \times \iota_S(b)$ in $P_S \times P_S$ that is transverse to the evaluation map from $\overline{\mathcal{M}}_{0,2}^\nu(P_S, J_S, \sigma_S)$, its image in

$P \times P$ will be transverse to the evaluation map from $\overline{\mathcal{M}}_{0,2}^\nu(P, J, \sigma)$ because $proj$ is a submersion at L_0 . Moreover, by [McDuff (2000)], we may suppose that α and β lie in distinct fibers of the projection $P_S \rightarrow S^2$. Let x_a, x_b be the corresponding points of \mathbf{CP}^n under the identification $S^2 = L_0$. Then every stable map that contributes to $n_P(\iota(a), \iota(b); \sigma)$ projects to an element of $\overline{\mathcal{M}}_{0,2}(\mathbf{CP}^n, j, [\mathbf{CP}^1])$ whose marked points map to the distinct points x_a, x_b . Since there is a unique line in \mathbf{CP}^n through two given points, in this case L_0 , every stable map that contributes to $n_P(\iota(a), \iota(b); \sigma)$ must project to L_0 and hence be contained in $\overline{\mathcal{M}}_{0,2}^\nu(P_S, J_S, \sigma_S)$. One can then check that

$$n_{P_S}(\iota_S(a), \iota_S(b); \sigma_S) = n_P(\iota(a), \iota(b); \sigma),$$

as claimed. The only delicate point here is to verify that the sign of each stable map on the left hand side is the same as the sign of the corresponding map on the right hand side. But this is also a consequence of the local triviality of the above projection map (see [Lalonde *et al.* (1998); Lalonde *et al.* (1999)] for more details).

The above argument generalizes easily to the case when B is a complex blowup of \mathbf{CP}^n .

Let B be a blowup of \mathbf{CP}^n along a *disjoint union* $Q = \coprod Q_i$ of complex submanifolds, each of complex codimension ≥ 2 . Then every Hamiltonian bundle B is c -split.

The above argument applies almost verbatim in the case when Q is a finite set of points. The top stratum of $\overline{\mathcal{M}}_{0,2}(B, j, [\mathbf{CP}^1])$ still consists of lines marked by two distinct points, and again all elements of this unperturbed moduli space are regular.

In the general case, there is a blow-down map $\psi : B \rightarrow \mathbf{CP}^n$ which is bijective over $\mathbf{CP}^n - Q$, and we can choose j on B so that the exceptional divisors $\psi^{-1}(Q)$ are j -holomorphic, and so that j is pulled back from the usual structure on \mathbf{CP}^n outside a small neighborhood of $\psi^{-1}(Q)$. Let L_0 be a complex line in $\mathbf{CP}^n - Q$. Then its pullback to B is still j -holomorphic. Hence, although the unperturbed moduli space $\overline{\mathcal{M}}_{0,2}(B, j, [\mathbf{CP}^1])$ may contain nonregular and hence ‘bad’ elements, its top stratum does contain an open set U_{L_0} consisting of marked lines near L_0 that are regular. Moreover, if we fix two points x_a, x_b on L_0 , every element of $\overline{\mathcal{M}}_{0,2}(B, j, [\mathbf{CP}^1])$ whose marked points map sufficiently close to x_a, x_b actually lies in this open set U_{L_0} . We can then regularize $\overline{\mathcal{M}}_{0,2}(B, j, [\mathbf{CP}^1])$ to a virtual moduli cycle that contains the open set U_{L_0} as part of its top stratum. More-

over, because the construction of the regularization is local with respect to $\overline{\mathcal{M}}_{0,2}(B, j, [\mathbf{CP}^1])$, this regularization $\overline{\mathcal{M}}_{0,2}^\nu(B, j, [\mathbf{CP}^1])$ will still have the property that each of its elements whose marked points map sufficiently close to x_a, x_b actually lies in this open set U_{L_0} .

We can now carry out the previous argument, choosing J on P to be fibered, and constructing ν to be compatible with the fibration on that part of $\overline{\mathcal{M}}_{0,2}(P, J, \sigma)$ that projects to U_{L_0} [Lalonde and McDuff (2002)]. Let $X = \#k\mathbf{CP}^2 \# \ell \overline{\mathcal{M}}^2$ be the connected sum of k copies of \mathbf{CP}^2 with ℓ copies of $\overline{\mathcal{M}}^2$. If one of k, ℓ is ≤ 1 then every bundle over X is c -split.

By reversing the orientation of X we can suppose that $k \leq 1$. The case $k = 1$ is covered in the previous proposition. When $k = 0$, pull the bundle back over the blowup of X at one point and then use the surjection lemma.

The previous proof can easily be generalized to the case of a symplectic base B that has a spherical 2-class A with Gromov–Witten invariant of the form $n_B(pt, pt, c_1, \dots, c_k; A)$ absolutely equal to 1¹⁹. Here c_1, \dots, c_k are arbitrary homology classes of B and we assume that $k \geq 0$. Again the idea is to construct the regularizations $\overline{\mathcal{M}}_{0,2+k}^\nu(P, J, \sigma)$ and $\overline{\mathcal{M}}_{0,2+k}^\nu(B, j, A)$ so that there is a projection from one to the other which is a fibration at least near the element of $\overline{\mathcal{M}}_{0,2+k}^\nu(B, j, A)$ that contributes to $n_B(pt, pt, c_1, \dots, c_k; A)$. Thus B could be the blowup of \mathbf{CP}^n along a symplectic submanifold Q that is disjoint from a complex line. One could also take similar blowups of products of projective spaces, or, more generally, of iterated fibrations of projective spaces. For example, if $B = \mathbf{CP}^m \times \mathbf{CP}^n$ with the standard complex structure then there is one complex line in the diagonal class $[\mathbf{CP}^1] + [\mathbf{CP}^1]$ passing through any two points and a cycle $H_1 \times H_2$, where H_i is the hyperplane class, and one could blow up along any symplectic submanifold that did not meet one such line.

It is also very likely that this argument can be extended to apply when all we know about B is that some Gromov–Witten invariant $n_B(pt, pt, c_1, \dots, c_k; A)$ is nonzero, for example, if B is a blowup of \mathbf{CP}^n along arbitrary symplectic submanifolds. There are two new problems here: (a) we must control the construction of $\overline{\mathcal{M}}_{0,2+k}^\nu(P, J, \sigma)$ in a neighborhood of all the curves that contribute to $n_B(pt, pt, c_1, \dots, c_k; A)$ and (b) we must make sure that the orientations are compatible so that curves in P projecting over different and noncancelling curves in B do not cancel each other in the global count of the Gromov–Witten invariant in P . Note

¹⁹By this we mean that for some generic j on B the relevant moduli space contains exactly one element, which moreover parametrizes an embedded curve in B .

that the bundles given by restricting P to the different curves counted in $n_B(pt, pt, c_1, \dots, c_k; A)$ are diffeomorphic, since, this being a homotopy theoretic question, we can always replace X by the simply connected space $X/(X_1)$ in which these curves are homotopic [McDuff (2000)].

4.12.2.5 Homotopy Reasons for Splitting

In this section we discuss c -splitting in a homotopy-theoretic context. Recall that a c -Hamiltonian bundle is a smooth bundle $P \rightarrow B$ together with a class $a \in H^2(P)$ whose restriction a_M to the fiber M is c -symplectic, i.e., $(a_M)^n \neq 0$ where $2n = \dim(M)$. Further a closed manifold M is said to satisfy the hard Lefschetz condition with respect to the class $a_M \in H^2(M, \mathbb{R})$ if the following maps are isomorphisms,

$$\cup(a_M)^k : H^{n-k}(M, \mathbb{R}) \rightarrow H^{n+k}(M, \mathbb{R}), \quad (1 \leq k \leq n).$$

In this case, elements in $H^{n-k}(M)$ that vanish when cupped with $(a_M)^{k+1}$ are called primitive, and the cohomology of M has an additive basis consisting of elements of the form $b \cup (a_M)^\ell$ where b is primitive and $\ell \geq 0$.²⁰

Let $M \rightarrow P \rightarrow B$ be a c -Hamiltonian bundle such that $\pi_1(B)$ acts trivially on $H^*(M, \mathbb{R})$. If in addition M satisfies the hard Lefschetz condition with respect to the c -symplectic class a_M , then the bundle c -splits [Blanchard (1956)].

The proof is by contradiction. Consider the Leray spectral sequence in cohomology and suppose that d_p is the first non zero differential. Then, $p \geq 2$ and the E_p term in the spectral sequence is isomorphic to the E_2 term and so can be identified with the tensor product $H^*(B) \otimes H^*(M)$. Because of the product structure on the spectral sequence, one of the differentials $d_p^{0,i}$ must be nonzero. So there is $b \in E_p^{0,i} \cong H^i(M)$ such that $d_p^{0,i}(b) \neq 0$. We may assume that b is primitive (since these elements together with a_M generate $H^*(M)$.) Then $b \cup a_M^{n-i} \neq 0$ but $b \cup a_M^{n-i+1} = 0$.

We can write

$$d_p(b) = \sum_j e_j \otimes f_j, \quad \text{where } e_j \in H^*(B) \quad \text{and} \quad f_j \in H^\ell(M) \quad (\text{with } \ell < i).$$

Hence $f_j \cup a_M^{n-i+1} \neq 0$ for all j by the Lefschetz property. Moreover, because the E_p term is a tensor product

$$(d_p(b)) \cup a_M^{n-i+1} = \sum_j e_j \otimes (f_j \cup a_M^{n-i+1}) \neq 0.$$

²⁰These manifolds are sometimes called ‘cohomologically Kähler manifolds.’

But this is impossible since this element is the image via d_p of the trivial element $b \cup a_M^{n-i+1}$ [Lalonde and McDuff (2002)].

Here is a related argument due to [Kedra (2000)]. Every Hamiltonian bundle with 4-dimensional fiber c -splits.

Consider the spectral sequence as above. We know that $d_2 = 0$ and $d_3 = 0$. Consider d_4 . We just have to check that $d_4^{0,3} = 0$ since $d_4^{0,i} = 0$ for $i = 1, 2$ for dimensional reasons, and $= 0$ for $i = 4$ since the top class survives.

Suppose $d_4(b) \neq 0$ for some $b \in H^3(M)$. Let $c \in H^1(M)$ be such that $b \cup c \neq 0$. Then $d_4(c) = 0$ and $d_4(b \cup c) = d_4(b) \cup c \neq 0$ since $d_4(b) \in H^4(B) \otimes H^0(M)$. But we need $d_4(b \cup c) = 0$ since the top class survives. So $d_4 = 0$ and then $d_k = 0$, $k > 4$ for reasons of dimension.

Here is an example of a c -Hamiltonian bundle over S^2 that is not c -split. This shows that c -splitting is a geometric rather than a topological (or homotopy-theoretic) property.

Observe that if S^1 acts on manifolds X, Y with fixed points p_X, p_Y then we can extend the S^1 action to the connected sum $X \# Y^{opp}$ at p_X, p_Y whenever the S^1 actions on the tangent spaces at p_X and p_Y are the same.²¹ Now let S^1 act on $X = S^2 \times S^2 \times S^2$ by the diagonal action in the first two spheres (and trivially on the third) and let the S^1 action on Y be the example constructed in [McDuff (1988)] of a non-Hamiltonian S^1 action that has fixed points. The fixed points in Y form a disjoint union of 2-tori and the S^1 action in the normal directions has index ± 1 . In other words, there is a fixed point p_Y in Y at which we can identify $T_{p_Y} Y$ with $\mathbf{C} \oplus \mathbf{C} \oplus \mathbf{C}$, where $\theta \in S^1$ acts by multiplication by $e^{i\theta}$ in the first factor, by multiplication by $e^{-i\theta}$ in the second and trivially in the third. Since there is a fixed point on X with the same local structure, the connected sum $Z = X \# Y^{opp}$ does support an S^1 -action. Moreover Z is a c -symplectic manifold. There are many possible choices of c -symplectic class: under the obvious identification of $H^2(Z)$ with $H^2(X) + H^2(Y)$ we will take the c -symplectic class on Z to be given by the class of the symplectic form on X .

Let $P_X \rightarrow S^2$, $P_Y \rightarrow S^2$ and $P_Z \rightarrow S^2$ be the corresponding bundles. Then P_Z can be thought of as the connect sum of P_X with P_Y along the sections corresponding to the fixed points. By analyzing the corresponding *Mayer-Vietoris sequence*, it is easy to check that the c -symplectic class on Z extends to P_Z . Further, the fact that the symplectic class in Y does not

²¹Here Y^{opp} denotes Y with the opposite orientation.

extend to P_Y implies that it does not extend to P_Z either. Hence $P_Z \rightarrow S^2$ is not c -split.

4.12.2.6 Action of the Homology of (M) on $H_*(M)$

The action $Ham(M) \times M \rightarrow M$ gives rise to maps

$$H_k(Ham(M)) \times H_*(M) \rightarrow H_{*+k}(M) : (\phi, Z) \mapsto \text{Tr}_\phi(Z).$$

We have the following result: These maps are trivial when $k \geq 1$.

To see this, let us first consider the action of a spherical element

$$\phi : S^n \rightarrow Ham(M).$$

It is not hard to check that the homomorphisms

$$\text{Tr}_\phi : H_k(M) \rightarrow H_{k+n}(M)$$

are precisely the connecting homomorphisms in the Wang sequence of the bundle $P_\phi \rightarrow S^{k+1}$ with clutching function ϕ , i.e., there is an exact sequence

$$\dots H_k(M) \xrightarrow{\text{Tr}_\phi} H_{k+n}(M) \rightarrow H_{k+n}(P) \xrightarrow{\cap[M]} H_{k-1}(M) \rightarrow \dots$$

Thus the fact that $P_\phi \rightarrow S^{k+1}$ is c -split immediately implies that the Tr_ϕ are trivial [Lalonde and McDuff (2002)].

Next recall that in a H -space the rational cohomology ring is generated by elements dual to the rational homotopy. It follows that there is a basis for $H_*(Ham(M))$ that is represented by cycles of the form

$$\phi_1 \times \dots \times \phi_k : S_1 \times \dots \times S_k \rightarrow Ham(M),$$

where the S_j s are spheres and one defines the product of maps by using the product structure in $Ham(M)$. Therefore it suffices to show that these product elements act trivially. However, if $a \in H_*(M)$ is represented by the cycle α , then $\text{Tr}_{S_k}(\alpha)$ is null-homologous, and so equals the boundary $\partial\beta$ of some chain β . Therefore,

$$\begin{aligned} \partial(\text{Tr}_{S_1 \times \dots \times S_{k-1}}(\beta)) &= \pm \text{Tr}_{S_1 \times \dots \times S_{k-1}}(\partial\beta) \\ &= \pm \text{Tr}_{S_1 \times \dots \times S_{k-1}}(\text{Tr}_{S_k}(\alpha)) = \text{Tr}_{S_1 \times \dots \times S_k}(\alpha). \end{aligned}$$

Hence $\text{Tr}_{S_1 \times \dots \times S_k}(a) = 0$.

Let $P \rightarrow B$ be a trivial symplectic bundle. Then any Hamiltonian automorphism $\Phi \in Ham(P, \pi)$ acts as the identity map on $H_*(P)$.

An element $\Phi \in \text{Ham}(P, \pi)$ is a map of the form

$$\Phi : B \times M \rightarrow B \times M : (b, x) \mapsto (b, \Phi_b(x)),$$

where $\Phi_b \in \text{Ham}(M)$ for all $b \in B$. Let us denote the induced map $B \times M \rightarrow M : (b, x) \mapsto \Phi_b(x)$ by α_Φ . The previous proposition implies that if B is a closed manifold of dimension > 0 , or, more generally, if it carries a fundamental cycle $[B]$ of degree > 0 ,

$$(\alpha_\Phi)_*([B] \otimes m) = \text{Tr}_{[B]}(m) = 0, \quad \text{for all } m \in H_*(M).$$

We can also think of $\Phi : B \times M \rightarrow B \times M$ as the composite

$$B \times M \xrightarrow{\text{diag}_B \times \text{Id}_M} B \times B \times M \xrightarrow{\text{Id}_B \times \alpha_\Phi} B \times M.$$

The diagonal class in $B \times B$ can be written as

$$[B] \otimes [pt] + \sum_{i \in I} b_i \otimes b'_i, \quad \text{where} \quad b_i, b'_i \in H_*(B) \quad \text{with} \quad \dim(b'_i) > 0.$$

Hence

$$\Phi_*([B] \otimes m) = [B] \otimes m + \sum_{i \in I} b_i \otimes \text{Tr}_{b'_i}(m) = [B] \otimes m.$$

More generally, given any class $b \in H_*(B)$, represent it by the image of the fundamental class $[X]$ of some polyhedron under a suitable map $X \rightarrow B$ and consider the pullback bundle $P_X \rightarrow X$. Since the class $\Phi_*([X] \otimes m)$ is represented by a cycle in $X \times M$ for any $m \in H_*(M)$, we can work out what it is by looking at the pullback of Φ to $X \times M$. The argument above then applies to show that $\Phi_*([X] \otimes m) = [X] \otimes m$ whenever b has degree > 0 . Thus $\Phi_* = \text{Id}$ on all cycles in $H_{*>0}(B) \otimes H_*(M)$. However, it clearly acts as the identity on $H_0(B) \otimes H_*(M)$ since the restriction of Φ to any fiber is isotopic to the identity.

We now show that there is a close relation between this question and the problem of c -splitting of bundles. Given an automorphism Φ of a symplectic bundle $M \rightarrow P \rightarrow B$ we define $P_\Phi = (P \times [0, 1])/\Phi$ to be the corresponding bundle over $B \times S^1$. If the original bundle and the automorphism are Hamiltonian, so is $P_\Phi \rightarrow B \times S^1$, though the associated bundle $P_\Phi \rightarrow B \times S^1 \rightarrow S^1$ over S^1 will not be, except in the trivial case when Φ is in the identity component of $\text{Ham}(P, \pi)$.

Assume that a given Hamiltonian bundle $M \rightarrow P \rightarrow B$ c -splits. Then a Hamiltonian automorphism $\Phi \in \text{Ham}(P, \pi)$ acts trivially (i.e., as the

identity) on $H_*(P)$ iff the corresponding Hamiltonian bundle $P_\Phi \rightarrow B \times S^1$ c -splits.

Clearly, the fibration $P \rightarrow B$ c -splits iff every basis of the \mathbf{Q} -vector space $H^*(M)$ can be extended to a set of classes in $H^*(P)$ that form a basis for a complement to the kernel of the restriction map. We will call such a set of classes a *Leray–Hirsch basis*. It corresponds to a choice of splitting isomorphism $H^*(P) \cong H^*(B) \otimes H^*(M)$. Now, the only obstruction to extending a Leray–Hirsch basis from P to P' is the nontriviality of the action of Φ on $H^*(P)$. This shows the “only if” part.

Conversely, suppose that P_Φ c -splits and let $e_j, j \in J$, be a Leray–Hirsch basis for $H^*(P_\Phi)$. Then $H^*(P_\Phi)$ has a basis of the form $e_j \cup \pi^*(b_i), e_j \cup \pi^*(b_i \times [dt])$ where b_i runs through a basis for $H^*(B)$ and $[dt]$ generates $H^1(S^1)$. Identify P with $P \times \{0\}$ in P_Φ and consider some cycle $Z \in H_*(P)$. Since the cycles $\Phi_*(Z)$ and Z are homologous in P_Φ , the classes $e_j \cup \pi^*(b_i)$ have equal values on $\Phi_*(Z)$ and Z . But the restriction of these classes to P forms a basis for $H^*(P)$. It follows that $[\Phi_*(Z)] = [Z]$ in $H_*(P)$.

Let $P \rightarrow B$ be a Hamiltonian bundle. Then the group $Ham(P, \pi)$ acts trivially on $H_*(P)$ if the base:

- (i) has dimension ≤ 2 , or
- (ii) is a product of spheres and projective spaces with no more than two S^1 factors, or
- (iii) is simply connected and has the property that all Hamiltonian bundles over B are c -split.

In all cases, the hypotheses imply that $P \rightarrow B$ c -splits. Therefore the previous proposition applies and (i), (ii) follow immediately from the splitting theorem. To prove (iii), suppose that B is a simply connected compact CW complex over which every Hamiltonian fiber bundle c -splits. Let $M \hookrightarrow P \rightarrow B \times S^1$ be any Hamiltonian bundle – in particular one of the form $P_\Phi \rightarrow B \times S^1$. Consider its pull-back P' by the projection map $B \times T^2 \rightarrow B \times S^1$. This is still a Hamiltonian bundle. To show that P c -splits, it is sufficient, to show that P' c -splits. $B \times T^2$ may be considered as a smooth compact Hamiltonian fibration $T^2 \hookrightarrow (B \times T^2) \rightarrow B$ with simply connected base, so P' c -splits since any Hamiltonian bundle over B or over T^2 c -splits [Lalonde and McDuff (2002)].

Finally, we prove the above statements about the automorphism groups of Hamiltonian structures.

We have to show that the following statements are equivalent for any $\Phi \in Symp_0(P, \pi)$:

- (i) Φ is isotopic to an element of $Ham(P, \pi)$;
- (ii) $\Phi^*({a}) = {a}$ for some Hamiltonian structure ${a}$ on P ;
- (iii) $\Phi^*({a}) = {a}$ for all Hamiltonian structures ${a}$ on P .

Recall that the relative homotopy groups $\pi_i(Symp(M), Ham(M))$ all vanish for $i > 1$. Using this together with the fact that $a \in H^2(P)$, we can reduce to the case when B is a closed oriented surface.

Let us prove this first in the case where $P \rightarrow B$ is trivial, so that Φ is a map $B \rightarrow Symp_0(M, \omega)$. Suppose that $\Phi^*(a) = a$ for some extension class a . By isotoping Φ if necessary, we can suppose that Φ takes the base point b_0 of B to the identity map. Then, the composite

$$\pi_1(B) \xrightarrow{\Phi_*} \pi_1(Symp_0(M)) \xrightarrow{Flux} H^1(M, \mathbb{R})$$

must vanish. Thus the restriction of $\Phi : B \rightarrow Symp_0(M)$ to the 1-skeleton of B homotops into $Ham(M)$. Since the relative homotopy groups $\pi_i(Symp(M), Ham(M))$ all vanish for $i > 1$, this implies that ϕ homotops to a map in $Ham(M)$, as required [Lalonde and McDuff (2002)].

Therefore, it remains to show that we can reduce the proof that (ii) implies (i) to the case when $P \rightarrow B$ is trivial. To this end, isotope Φ so that it is the identity map on all fibers M_b over some disc $D \subset B$. Since $P \rightarrow B$ is trivial over $X = B - D$, we can decompose $P \rightarrow B$ into the fiber connected sum of a trivial bundle P_B over B (where B is thought of as the space obtained from X by identifying its boundary to a point) and a nontrivial bundle Q over $S^2 = D/\partial D$. Further, this decomposition is compatible with Φ , which can be thought of as the fiber sum of some automorphism Φ_B of P_B together with the trivial automorphism of Q . Clearly, this reduces the proof that (ii) implies (i) to the case $\Phi_B : P_B \rightarrow P_B$ on trivial bundles, if we note that when Φ_B is the identity over some disc $D \subset B$, the isotopy between Φ and an element in $Ham(P_B)$ can be constructed so that it remains equal to the identity over D .

4.12.2.7 Cohomology of General Symplectic Bundles

Now we discuss some consequences for general symplectic bundles of the above results on Hamiltonian bundles. First, we prove the above proposition, which states that the boundary map ∂ in the rational homology Wang sequence of a symplectic bundle over S^2 has $\partial \circ \partial = 0$.

The map

$$\partial \circ \partial : H_k(M) \rightarrow H_{k+2}(M) \quad \text{is given by} \quad a \mapsto \Psi_*([T^2] \otimes a),$$

where $\Psi : T^2 \times M \rightarrow M : (s, t, x) \mapsto \phi_s \phi_t(x)$.

Let $b(s, t) = \phi_{s+t}^{-1} \phi_s \phi_t$. The map $(s, t) \mapsto b(s, t)$ factors through

$$T^2 \longrightarrow S^2 = T^2 / \{s = 0\} \cup \{t = 0\}.$$

Let $Z \rightarrow M$ represent a k -cycle. We have a map

$$T^2 \times Z \xrightarrow{A_1} S^1 \times S^2 \times Z \xrightarrow{A_2} M, \quad \text{given by}$$

$$(s, t, z) \mapsto (\phi_{s+t}, b(s, t), z) \mapsto \phi_{s+t} b(s, t) z = \phi_s \phi_t(z),$$

and want to calculate

$$\int_{T^2 \times Z} A_1^* A_2^*(\alpha) = \int_{(A_1)_*[T^2 \times Z]} A_2^*(\alpha),$$

for some $k+2$ -form α on M . But $(A_1)_*[T^2 \times Z] \in H_2(S^2) \otimes H_k(Z)$.²² Now observe that $A_2^*(\alpha)$ vanishes on $H_2(S^2) \otimes H_k(Z)$.

The previous lemma is trivially true for any smooth (not necessarily symplectic) bundle over S^2 that extends to \mathbf{CP}^2 . For the differential d_2 in the *Leray cohomology spectral sequence* can be written as

$$d_2(a) = \partial(a) \cup u \in E_2^{2, q-1},$$

where $a \in H^q(M) \equiv E_2^{0, q}$ and u generates $H^2(\mathbf{CP}^2) \equiv E_2^{2, 0}$. Hence

$$0 = d_2(d_2(a)) = d_2(\partial(a) \cup u) = d_2(\partial(a)) \cup u = \partial(\partial(a)) \otimes u^2.$$

If $\pi : P \rightarrow B$ is any symplectic bundle over a simply connected base, then $d_3 \equiv 0$.

As above we can reduce to the case when B is a wedge of S^2 s and S^3 s. The differential d_3 is then given by restricting to the bundle over $\vee S^3$. Since this is Hamiltonian, $d_3 \equiv 0$.

The next lemma describes the *Wang differential* $\partial = \partial_\phi$ in the case of a symplectic loop ϕ with nontrivial image in $H_1(M)$: Suppose that ϕ is a symplectic loop such that $[\phi_t(x)] \neq 0$ in $H_1(M)$. Then $\text{Ker } \partial = \text{Im } \partial$, where $\partial = \partial_\phi : H^k(M) \rightarrow H^{k-1}(M)$ is the corresponding Wang differential.

²²Note that there is no component in $H_3(S^1 \times S^2) \otimes H_{k-1}(Z)$ since $A_1 = \text{Id}$ on the Z factor.

Let $\alpha \in H^1(M)$ be such that $\alpha([\phi_t(x)]) = 1$. So $\partial\alpha = 1$. Then, for every $\beta \in \text{Ker } \partial$, $\partial(\alpha \cup \beta) = \beta$. As $\partial \circ \partial = 0$, this means that [Lalonde *et al.* (1998); Lalonde *et al.* (1999); Lalonde and McDuff (2002)]

$$\text{Ker } \partial \subset \text{Im } \partial \quad \implies \quad \text{Ker } \partial = \text{Im } \partial.$$

Moreover, the map

$$\alpha \cup : H^k(M) \rightarrow H^{k+1}(M)$$

is injective on $\text{Ker } \partial$ and $H^*(M)$ decomposes as the direct sum $\text{Ker } \partial \oplus (\alpha \cup \text{Ker } \partial)$.

The above corollary claims that for a symplectic loop ϕ ,

$$\text{Ker } \partial = \text{Im } \partial \quad \text{iff} \quad [\phi_t(x)] \neq 0 \quad \text{in } H_1(M).$$

Since $1 \in H^0(M)$ is in $\text{Ker } \partial$ it must equal $\partial(\alpha)$ for some $\alpha \in H^1(M)$. This means that $\alpha([\phi_t(x)]) \neq 0$ so that $[\phi_t(x)] \neq 0$. Note that the only place that the *symplectic condition* enters in the above proof is in the claim that $\partial \circ \partial = 0$. Since this is always true when the loop comes from a circle action, this lemma holds for all, not necessarily symplectic, circle actions. In this case, we can interpret the result topologically. For the hypothesis $[\phi_t(x)] \neq 0$ in $H_1(X)$ implies that the action has no fixed points, so that the quotient M/S^1 is an orbifold with cohomology isomorphic to $\text{ker } \partial$. Thus, the argument shows that M has the same cohomology as the product $(M/S^1) \times S^1$.

4.13 Clifford Algebras, Spinors and Penrose Twistors

4.13.1 Clifford Algebras and Modules

In this subsection, mainly following [Yang (1995)], we provide the general theory of Clifford algebra and subsequently consider its special 4D case.

Let V be a vector space over \mathbb{R} with a quadratic form Q on it. The *Clifford algebra* of (V, Q) , denoted by $C(V, Q)$, is the algebra over \mathbb{R} generated by V with the relations

$$v_1 \cdot v_2 + v_2 \cdot v_1 = -2Q(v_1, v_2), \quad (\text{for all } v_1, v_2 \in V).$$

Since Q is symmetric, we have $v^2 = -Q(v)$ for all $v \in V$.

For fixed Q , we may abbreviate $C(V, Q)$ and $Q(v_1, v_2)$ into $C(V)$ and (v_1, v_2) respectively.

It is a basic fact in algebra that the Clifford algebra is the unique (up to isomorphism) solution to the following universal problem.

If A is an algebra and $c : V \longrightarrow A$ is a linear map satisfying

$$c(v_2)c(v_1) + c(v_1)c(v_2) = -2Q(v_1, v_2), \quad (\text{for all } v_1, v_2 \in V),$$

then there is a unique algebra homomorphism from $C(V, Q)$ to A extending the map c .

The Clifford algebra may be realized as the quotient $T(V)/\mathcal{I}_Q$ where

$$T(V) = \bigoplus_{k=1}^{\infty} T^k(V)$$

is the *tensor algebra* of V with $T^k(V)$ generated by

$$\{v_1 \otimes v_2 \otimes \cdots \otimes v_k \mid v_1, v_2, \dots, v_k \in V\}$$

and \mathcal{I}_Q is generated by

$$\{v_1 \otimes v_2 + v_2 \otimes v_1 + 2Q(v_1, v_2) \mid v_1, v_2 \in V\}$$

The tensor algebra $T(V)$ has a \mathbb{Z}_2 -grading obtained from the natural \mathbb{N} -grading after reduction mod 2:

$$\begin{aligned} T(V) &= T^+(V) + T^-(V), \quad \text{where} \\ T^+(V) &= \mathbb{R} \oplus T^2(V) \oplus T^4(V) \oplus \cdots \oplus T^{2k}(V) \oplus \cdots, \\ T^-(V) &= V \oplus T^3(V) \oplus T^5(V) \oplus \cdots \oplus T^{2k+1}(V) \oplus \cdots \end{aligned}$$

Therefore it forms a *super-algebra*.

Similarly, for $k = 0, 1, 2, 3, \dots$, let

$$\begin{aligned} C^k(V) &= T^k(V)/\mathcal{I}_Q, \quad \text{and let} \\ C^+(V) &= \mathbb{R} \oplus C^2(V) \oplus C^4(V) \oplus \cdots \oplus C^{2k}(V) \oplus \cdots, \\ C^-(V) &= V \oplus C^3(V) \oplus C^5(V) \oplus \cdots \oplus C^{2k+1}(V) \oplus \cdots \end{aligned}$$

Since the ideal \mathcal{I}_Q is generated by elements from the evenly graded subalgebra $T^+(V)$, $C(V)$ is itself a superalgebra and we have the grading

$$C(V) = C^+(V) + C^-(V).$$

Let E be a module over \mathbb{R} or \mathbb{C} which is \mathbb{Z}_2 -graded,

$$E = E^+ \oplus E^-.$$

E is called a *Clifford module* over a Clifford algebra $C(V)$ if there is a *Clifford action*

$$\begin{aligned} C(V) \times E &\xrightarrow{\cdot} E \\ (a, e) &\longmapsto a \cdot e \end{aligned}$$

or equivalently, an algebra homomorphism

$$\begin{aligned} C(V) &\xrightarrow{\cdot} \text{End}(E) \\ a &\longmapsto c(a) \end{aligned} \quad \text{with} \quad c(a)(e) = a \cdot e,$$

which is *even* with respect to this grading:

$$C^+(V) \cdot E^\pm \subset E^\pm, \quad C^-(V) \cdot E^\mp \subset E^\mp.$$

Let $O(V, Q)$ be the group of linear transformations of V which preserve Q . That means for all $\phi \in O(V, Q)$ and $v_1, v_2 \in V$,

$$Q(\phi v_1, \phi v_2) = Q(v_1, v_2).$$

The action of $O(V, Q)$ on generators of $T(V)$ is defined by

$$\phi(v_1 \otimes v_2 \otimes \cdots \otimes v_k) = \sum_{i=1}^k v_1 \otimes \cdots \otimes \phi(v_i) \otimes \cdots \otimes v_k$$

and extends to the whole $T(V)$ linearly.

\mathcal{I}_Q is invariant under the action of $O(V, Q)$. Hence $C(V, Q)$ carries a natural action of $O(V, Q)$.

Let $*$: $a \mapsto a^*$ be the anti-automorphism of $T(V)$ induced by $v \mapsto -v$ on T , and satisfies

$$(a_1 a_2)^* = a_2^* a_1^*.$$

Hence,

$$(v_1 v_2 \cdots v_k)^* = \begin{cases} (v_k v_{k-1} \cdots v_1), & \text{if } k \text{ is even,} \\ -(v_k v_{k-1} \cdots v_1), & \text{if } k \text{ is odd.} \end{cases}$$

Since

$$\begin{aligned} &v_1^* \otimes v_2^* + v_2^* \otimes v_1^* + 2Q(v_1^*, v_2^*) \\ &= (-v_1) \otimes (-v_2) + (-v_2) \otimes (-v_1) + 2Q(-v_1, -v_2) \\ &= v_1 \otimes v_2 + v_2 \otimes v_1 + 2Q(v_1, v_2), \end{aligned}$$

\mathcal{I}_Q is invariant under $*$. So it induces an anti-isomorphism $a \mapsto a^*$ of $C(V)$.

If Q is a positive-definite quadratic form, then a Clifford module E of $C(V)$ with an inner product is said to be *self-adjoint* if $c(a^*) = c(a)^*$, where $c(v)$ denote the action of $v \in V$ on a Clifford module of $C(V)$ (which may be $C(V)$ itself).

The inner product on E must be $C(V)$ – *invariant*:

$$(c(a)e_1, c(a)e_2) = (e_1, e_2), \quad (\text{for all } a \in C(V), e_1, e_2 \in E).$$

Hence for a self-dual module E ,

$$(c(a)e_1, e_2) = (e_1, c(a)^*e_2) = (e_1, c(a^*)e_2).$$

Especially, we have

$$(c(v)e_1, e_2) = (e_1, c(v^*)e_2) = (e_1, -c(v)e_2), \quad (\text{for all } v \in V, e_1, e_2 \in E).$$

That means $c(v)$ is *skew-adjoint* for all $v \in V$.

Let E be a \mathbb{Z}_2 –graded Clifford module over the Clifford algebra $C(V)$. We denote by $\text{End}_{C(V)}(E)$ the algebra of homomorphisms of E supercommuting with the action of $C(V)$.

4.13.1.1 The Exterior Algebra

The first interesting example of Clifford module is the exterior algebra.

The *exterior algebra* ΛV of a vector space V is defined to be $T(V)/\mathcal{I}_*$, where \mathcal{I}_* is the ideal generated by elements of the form

$$(v_1 \otimes \cdots \otimes v_i \otimes v_{i+1} \otimes \cdots \otimes v_k) + (v_1 \otimes \cdots \otimes v_{i+1} \otimes v_i \otimes \cdots \otimes v_k),$$

or equivalently,

$$\Lambda V = \bigoplus_{k=1}^{\infty} \Lambda^k V, \quad \text{where} \quad \Lambda^k V = T^k(V)/\mathcal{I}_*.$$

Let

$$\epsilon : V \longrightarrow \text{Hom}(\Lambda^k V, \Lambda^{k+1} V)$$

be the action of V on ΛV by *exterior product*, i.e., for all $v \in V$,

$$\begin{aligned} \epsilon(v) : \Lambda^k V &\longrightarrow \Lambda^{k+1} V \\ w &\longmapsto v \wedge w. \end{aligned}$$

Explicitly,

$$\epsilon(v)(v_1 \wedge \cdots \wedge v_k) = v \wedge v_1 \wedge \cdots \wedge v_k.$$

Let

$$\iota : V \longrightarrow \text{Hom}(\Lambda^k V, \Lambda^{k-1} V)$$

be the action of V on ΛV by *interior product* (or *contraction*), i.e., for all $v \in V$,

$$\begin{aligned} \iota(v) : \Lambda^k V &\longrightarrow \Lambda^{k-1} V \\ w &\longmapsto Q(v, w). \end{aligned}$$

Explicitly,

$$\iota(v)(v_1 \wedge \cdots \wedge v_k) = \sum_{i=1}^k (-1)^{i-1} Q(v, v_i) v_1 \wedge \cdots \wedge \widehat{v_i} \wedge \cdots \wedge v_k.$$

The Clifford action of $v \in V$ on $w \in \Lambda V$ is given by

$$v \cdot_c w = c(v)w = \epsilon(v)w - \iota(v)w.$$

For any v, w in V , we have

$$\epsilon(v)\iota(w) + \iota(w)\epsilon(v) = Q(v, w).$$

The action $c : V \longrightarrow \text{End}(\Lambda V)$ extends to an action of the Clifford algebra $C(V)$ on ΛV .

The *symbol map* $\sigma : C(V) \longrightarrow \Lambda V$ is defined by $\sigma(a) = c(a)1_{\Lambda V}$, where $1_{\Lambda V} \in \Lambda^0 V$ is the identity in the exterior algebra ΛV .

If $1_{C(V)}$ denotes the identity in $C(V)$, then $\sigma(1_{C(V)})$ is the identity $1_{\text{End}(\Lambda V)}$ in $\text{End}(\Lambda V)$.

The Clifford algebra $C(V)$ is isomorphic to the tensor algebra ΛV and is therefore a $2^{\dim V}$ dimensional vector space with generators

$$\{(c_1)^{n_1}(c_2)^{n_2} \cdots (c_{\dim V})^{n_{\dim V}} \mid (n_1, n_2, \dots, n_{\dim V}) \in \{0, 1\}^{\dim V}\}.$$

If we consider $C(V)$ and ΛV as \mathbb{Z}_2 -graded $O(V)$ -modules, then σ and preserve the \mathbb{Z}_2 -graded and the $O(V)$ action. Hence they are isomorphisms of \mathbb{Z}_2 -graded $O(V)$ -modules.

There is a natural increasing *filtration*

$$C_0(V) \subseteq C_1(V) \subseteq \cdots \subseteq C_k(V) \subseteq \cdots \subseteq \bigcup_{i=0}^{\infty} C_i(V) = C(V),$$

where

$$C_i(V) = C^0(V) \oplus C^1(V) \oplus C^2(V) \oplus \dots \oplus C^i(V)$$

and $C^0(V) = \mathbb{R}$. It follows that

$$C_i(V) = \text{span}\{v_1 \dots v_k \mid v_j \in V \hookrightarrow C(V) \text{ for } j = 1, \dots, k \leq i\}.$$

The Clifford algebra $C(V)$ with this filtration is called the *associated graded algebra* of $C(V)$ and is denoted by $\text{gr } C(V)$. The i th grading of $\text{gr } C(V)$ is denoted by $\text{gr}_i C(V)$.

The associated graded algebra $\text{gr } C(V)$ is naturally isomorphic to the exterior algebra ΛV , where the isomorphism is given by sending $\text{gr}_i(v_1 \dots v_i) \in \text{gr}_i C(V)$ to $v_1 \wedge \dots \wedge v_i \in \Lambda^i V$. The symbol map σ extends the symbol map

$$\sigma_i : C_i(V) \longrightarrow \text{gr}_i C(V) \cong \Lambda^i V,$$

in the sense that if $a \in C_i(V)$, then $\sigma(a)_{[i]} = \sigma_i(a)$. The filtration $C_i(V)$ may be written as

$$C_i(V) = \sum_{j=0}^i q(\Lambda^j V).$$

Hence the Clifford algebra $C(V)$ may be identified with the exterior algebra ΛV with a *twisted*, or *quantized* multiplication $\alpha \cdot_Q \beta$.

If $v \in V \hookrightarrow C(V)$ and $a \in C^+(V)$, then $\sigma([v, a]) = -2\iota(v)\sigma(a)$.

The space $C^2(V) = q(\Lambda^2 V)$ is a Lie subalgebra of $C(V)$, where the Lie bracket is just the commutator in $C(V)$. It is isomorphic to the Lie algebra (V) , under the map $\tau : C^2(V) \longrightarrow \mathfrak{so}(V)$, where any $a \in C^2(V)$ is mapped into $\tau(a)$ which acts on any $v \in V \cong C^1(V)$ by the *adjoint action*: $\tau(a)v = [a, v]$. Here the bracket is the bracket of the *Lie super-algebra* $C(V)$, i.e.,

$$[a_1, a_2] = a_1 a_2 - (-1)^{|a_1||a_2|} a_2 a_1, \quad (\text{for } a_1 \in C^{|a_1|}(V), a_2 \in C^{|a_2|}(V)).$$

It satisfies the following *Axioms of Lie super-algebra*:

Supercommutativity $[a_1, a_2] + (-1)^{|a_1||a_2|} [a_2, a_1] = 0$, and
Jacobi's identity

$$[a_1, [a_2, a_3]] = [[a_1, a_2], a_3] + (-1)^{|a_1||a_2|} [a_2, [a_1, a_3]].$$

We usually identify $A \in \mathfrak{so}(V)$ with

$$\sum_{i < j} (A e_i, e_j) e_i \wedge e_j \in \Lambda^2 V, \quad \text{so that we have} \quad q(A) = \sum_{i < j} (A e_i, e_j) c_i c_j,$$

which is *twice* of $\tau^{-1}(A)$.

4.13.1.2 The Spin Group

For any a in the Lie algebra $C(V)$, we may form its *exponential* in $C(V)$ by

$$\exp(a) = 1_{C(V)} + a + \frac{1}{2} a^2 + \frac{1}{3!} a^3 + \dots + \frac{1}{n!} a^n + \dots,$$

which is an element in the associated Lie group of $C(V)$.

For any v_1, v_2 in $V \hookrightarrow C(V)$ satisfying

$$Q(v_1, v_1) = Q(v_2, v_2) = 1, \quad Q(v_1, v_2) = 0,$$

we have the following formula

$$\exp[t(v_1 v_2)] = (\cos t) 1_{C(V)} + (\sin t) v_1 v_2,$$

where $t \in \mathbb{R}$. In fact, t is well-defined mod 2π . Consequently, this formula is satisfied for some vectors in V whenever $\dim V > 1$.

Let the *Spin group* of the vector space V be the Lie group associated to the Lie subalgebra $C^2(V)$ of the Clifford algebra $C(V)$, i.e., $\text{Spin}(V) = \exp C^2(V)$.

The adjoint action τ of the Lie algebra $C^2(V)$ on V may be exponentiated to an orthogonal action of *conjugation* which is denoted by $\underline{\tau}$. Explicitly, for $g \in \text{Spin}(V)$ and $v \in V$, there is a fundamental relation: $\underline{\tau}(g)v = gv g^{-1}$.

Indeed, writing $g = \exp(a)$ for some $a \in C^2(V)$, then

$$[a, v] = \tau(a)v \quad \text{implies} \quad \exp(a)v(\exp(a))^{-1} = \exp(\tau(a))v.$$

If $\dim V > 1$, then the homomorphism $\underline{\tau} : \text{Spin}(V) \longrightarrow SO(V)$ is a double covering.

4.13.1.3 4D Case

Now consider the most interesting case when $V \cong \mathbb{R}^4$. Fix a basis $\{e_1, e_2, e_3, e_4\}$ which is orthonormal with respect to the fixed quadratic

form Q . Then any vector $v \in V$ may be written as $v = v^k e_k$ and the Clifford algebra is

$$C(\mathbb{R}^4) = \text{span}\{c_1^{n_1} c_2^{n_2} c_3^{n_3} c_4^{n_4} \mid n_i \in \{0, 1\}\}, \quad \text{where} \quad c_i = q(e_i).$$

For convenience, we may denote c_i by e_i without ambiguity. Especially,

$$C^2(\mathbb{R}^4) = \text{span}\{e_1 e_2, e_1 e_3, e_1 e_4, e_2 e_3, e_2 e_4, e_3 e_4\}$$

Now consider the isomorphism $\tau : C^2(\mathbb{R}^4) \longrightarrow \mathfrak{so}(4) = \mathfrak{so}(\mathbb{R}^4)$. And the corresponding $\tau : Spin(4) \longrightarrow \mathfrak{so}(4)$. We have

$$\begin{aligned} \tau(e_i e_j) \cdot v &= [e_i e_j, \sum_{k=1}^4 v^k e_k] = \sum_{k=1}^4 v^k [e_i e_j, e_k] = \sum_{k=1}^4 v^k (e_i e_j e_k - e_k e_i e_j) \\ &= \sum_{k=1}^4 v^k (e_i e_j e_k + e_i e_k e_j - e_i e_k e_j - e_k e_i e_j) \\ &= \sum_{k=1}^4 v^k (-2Q(e_j, e_k) e_i + 2Q(e_k, e_i) e_j) = -2v^j e_i + 2v^i e_j. \end{aligned}$$

Hence $\tau(e_i e_j)$ corresponds to $2 \cdot m(i, j) \in \mathfrak{so}(4)$ where $m(i, j) = (m(i, j)_{\alpha\beta})$ is a matrix with entries

$$m(i, j)_{\alpha\beta} = \begin{cases} 1, & \text{if } \alpha = j \text{ and } \beta = i, \\ -1, & \text{if } \alpha = i \text{ and } \beta = j, \\ 0, & \text{otherwise.} \end{cases}$$

Explicitly, we have

$$\tau(e_1 e_2) = 2 \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \tau(e_1 e_3) = 2 \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\tau(e_1 e_4) = 2 \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \quad \tau(e_2 e_3) = 2 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\tau(e_2 e_4) = 2 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \quad \tau(e_3 e_4) = 2 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Notice that for $t \in [0, 2\pi)$, $\tau(\frac{t}{2} e_i e_j)$ corresponds to the matrix $t \cdot m(i, j)$ and

$$(t \cdot m(i, j)_{\alpha\beta})^2 = -t^2 \cdot \Delta(i, j),$$

$$(t \cdot m(i, j)_{\alpha\beta})^4 = t^4 \cdot \Delta(i, j),$$

where $\Delta(i, j)$ is a matrix with

$$\Delta(i, j)_{\alpha\beta} = \begin{cases} 1, & \text{if } \alpha = \beta = i \text{ or } \alpha = \beta = j, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, we have

$$\exp \tau \left(\frac{t}{2} e_i e_j \right) = (\cos t - 1) \cdot \Delta(i, j) + 1_{SO(4)} + (\sin t) m(i, j).$$

Since we have the commutative relation $\underline{\tau} \cdot \exp = \exp \cdot \tau$, it follows that

$$\underline{\tau} \exp(t e_i e_j) = \exp \tau(t e_i e_j) = (\cos 2t - 1) \cdot \Delta(i, j) + 1_{SO(4)} + (\sin 2t) m(i, j).$$

Explicitly, we have

$$\underline{\tau}(\exp(t e_1 e_2)) = \begin{pmatrix} \cos 2t - \sin 2t & 0 & 0 \\ \sin 2t & \cos 2t & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\underline{\tau}(\exp(t e_1 e_3)) = \begin{pmatrix} \cos 2t & 0 - \sin 2t & 0 \\ 0 & 1 & 0 & 0 \\ \sin 2t & 0 & \cos 2t & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\underline{\tau}(\exp(t e_1 e_4)) = \begin{pmatrix} \cos 2t & 0 & 0 - \sin 2t \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \sin 2t & 0 & 0 & \cos 2t \end{pmatrix}$$

$$\mathcal{T}(\exp(t e_2 e_3)) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2t & -\sin 2t & 0 \\ 0 & \sin 2t & \cos 2t & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\mathcal{T}(\exp(t e_2 e_4)) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2t & 0 & -\sin 2t \\ 0 & 0 & 1 & 0 \\ 0 & \sin 2t & 0 & \cos 2t \end{pmatrix}$$

$$\mathcal{T}(\exp(t e_3 e_4)) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos 2t & -\sin 2t \\ 0 & 0 & \sin 2t & \cos 2t \end{pmatrix}.$$

4.13.2 Spinors

In this subsection, mainly following [Yang (1995)], we define spinors and the spinor representation of the Clifford algebras.

4.13.2.1 Basic Properties

Let $\{e_i\}_{i=1,\dots,\dim V}$ be an *oriented*, orthonormal basis of V , that means there is a preferred ordering of the basis elements modulo even permutations. The *chirality operator*²³ is

$$\Gamma = i^p e_1 \dots e_{\dim V} \in C(V) \otimes \mathbb{C}, \quad \text{where}$$

$$p = \left\lfloor \frac{\dim V + 1}{2} \right\rfloor = \begin{cases} \frac{\dim V}{2} & \text{if } \dim V \text{ is even,} \\ \frac{\dim V + 1}{2} & \text{if } \dim V \text{ is odd.} \end{cases}$$

The Chirality operator satisfies: $\Gamma^2 = 1_{C(V)}$, and it *super-anticommutes* with elements $v \in V$, i.e., $\Gamma v + (-1)^{\dim V} v \Gamma = 0$.

For $\dim V$ odd, Γ is in the center of $C(V) \otimes \mathbb{C}$.

²³Also known as the *volume element* or the *complex unit*. Physicists also denote it as γ^5 for the four dimensional case.

For $\dim V$ even, every complex Clifford module E has a \mathbb{Z}_2 -grading defined by the ± 1 eigen-spaces of the chirality operator:

$$E^\pm = \{v \in E \mid \Gamma v = \pm v\}.$$

Especially, for $\dim V \equiv 0 \pmod{4}$, $\Gamma \in C(V)$, so in this case, the real Clifford modules are also \mathbb{Z}_2 -graded.

A *polarization* of a complex vector space $V \otimes \mathbb{C}$ is a subspace $P \subset V \otimes \mathbb{C}$ which is *isotropic*, i.e.,

$$Q(v, v) = 0, \quad (\text{for all } v \in P),$$

and we have a splitting

$$V \otimes \mathbb{C} = \mathbb{P} \oplus \bar{\mathbb{P}}.$$

Here the quadratic form Q extends from V to $V \otimes \mathbb{C}$ *complex linearly*, i.e.,

$$Q(a + ib, c + id) = Q(a, c) + i(Q(a, d) + Q(b, c)) - Q(b, d)$$

A polarization is called *oriented*, if there is an oriented orthonormal basis $\{e_i\}$ of V , such that P is spanned by the vectors

$$\{w_i = \frac{(e_{2i-1} - ie_{2i})}{\sqrt{2}} \mid 1 \leq i \leq \frac{\dim V}{2}\},$$

and therefore the complement \bar{P} is spanned by the vectors

$$\{\bar{w}_i = \frac{(e_{2i-1} + ie_{2i})}{\sqrt{2}} \mid 1 \leq i \leq \frac{\dim V}{2}\}.$$

The basis $\{w_i\}_{i=1, \dots, \frac{\dim V}{2}}$ and the corresponding conjugate $\{\bar{w}_i\}_{i=1, \dots, \frac{\dim V}{2}}$ satisfy the following equations: For $1 \leq i \leq \frac{\dim V}{2}$,

$$w_i w_i = \bar{w}_i \bar{w}_i = 0, \quad w_i \bar{w}_i + \bar{w}_i w_i = -2.$$

For $1 \leq i \neq j \leq \frac{\dim V}{2}$,

$$w_i w_j = -w_j w_i, \quad w_i \bar{w}_j = -\bar{w}_j w_i, \quad \bar{w}_i \bar{w}_j = -\bar{w}_j \bar{w}_i.$$

If V is an even-dimensional oriented Euclidean vector space, then there is a unique \mathbb{Z}_2 -graded Clifford module: $S = S^+ \oplus S^-$ called the *spinor module*, such that $C(V) \otimes \mathbb{C} \cong \text{End}(S)$. Elements in S^+ and S^- are called *positive* and *negative spinors* respectively.

In particular, we have

$$\dim_{\mathbb{C}}(S) = 2^{\left(\frac{\dim V}{2}\right)}, \quad \text{and} \quad \dim_{\mathbb{C}}(S^+) = \dim_{\mathbb{C}}(S_-) = 2^{\left(\frac{\dim V}{2}-1\right)}.$$

4.13.2.2 $4D$ Case

Now we look at the spinors in four dimensional vector space.

Consider an oriented orthonormal basis $\{e_1, e_2, e_3, e_4\}$ of the Euclidean vector space \mathbb{R}^4 . Then the standard oriented polarization P of $\mathbb{C}^4 \cong \mathbb{R}^4 \otimes \mathbb{C}$ is generated by

$$\left\{ w_1 = \frac{e_1 - i e_2}{\sqrt{2}}, \quad w_2 = \frac{e_3 - i e_4}{\sqrt{2}} \right\}, \quad \text{and we have}$$

$$S^+ = \text{span}_{\mathbb{C}} \langle 1_{\wedge P}, w_1 \wedge w_2 \rangle, \quad S^- = \text{span}_{\mathbb{C}} \langle w_1, w_2 \rangle.$$

Together, we have the following standard basis of $S = S^+ \oplus S^-$:

$$\{ 1_{\wedge P}, w_1 \wedge w_2, w_1, w_2 \}.$$

Under this basis, any spinor $s \in S$ may be written as a column vector with components

$$s = \begin{pmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \end{pmatrix}$$

and we have a splitting

$$s = s^+ \oplus s^- = \begin{pmatrix} s_1^+ \\ s_2^+ \end{pmatrix} \oplus \begin{pmatrix} s_1^- \\ s_2^- \end{pmatrix} = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \oplus \begin{pmatrix} s_3 \\ s_4 \end{pmatrix}.$$

Since $S \cong \mathbb{C}^4$, we have a representation: $C(V) \otimes \mathbb{C} \hookrightarrow \text{End}(\mathbb{C}^4)$. To find out the exact correspondences, let's consider the Clifford action of w_i

and \overline{w}_i on S :

$$c(w_1) : \begin{cases} 1_{\wedge P} \\ w_1 \wedge w_2 \\ w_1 \\ w_2 \end{cases} \mapsto \begin{cases} \sqrt{2} w_1 \\ 0 \\ 0 \\ \sqrt{2} w_1 \wedge w_2 \end{cases}$$

$$c(w_2) : \begin{cases} 1_{\wedge P} \\ w_1 \wedge w_2 \\ w_1 \\ w_2 \end{cases} \mapsto \begin{cases} \sqrt{2} w_2 \\ 0 \\ -\sqrt{2} w_1 \wedge w_2 \\ 0 \end{cases}$$

$$c(\overline{w}_1) : \begin{cases} 1_{\wedge P} \\ w_1 \wedge w_2 \\ w_1 \\ w_2 \end{cases} \mapsto \begin{cases} 0 \\ -\sqrt{2} w_2 \\ -\sqrt{2} 1_{\wedge P} \\ 0 \end{cases}$$

$$c(\overline{w}_2) : \begin{cases} 1_{\wedge P} \\ w_1 \wedge w_2 \\ w_1 \\ w_2 \end{cases} \mapsto \begin{cases} 0 \\ \sqrt{2} w_1 \\ 0 \\ -\sqrt{2} 1_{\wedge P} \end{cases}.$$

Therefore we have the following correspondence:

$$w_1 = \sqrt{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$w_2 = \sqrt{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

$$\bar{w}_1 = \sqrt{2} \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}$$

$$\bar{w}_2 = \sqrt{2} \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Now, by the relations

$$e_{2i-1} = \frac{(w_i + \bar{w}_i)}{\sqrt{2}}, \quad e_{2i} = \frac{(-w_i + \bar{w}_i)}{\sqrt{2}\mathbf{i}},$$

we can deduce the corresponding matrices:

$$e_1 = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}$$

$$e_2 = \begin{pmatrix} 0 & 0 & i & 0 \\ 0 & 0 & 0 & i \\ i & 0 & 0 & 0 \\ 0 & i & 0 & 0 \end{pmatrix}$$

$$e_3 = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

$$e_4 = \begin{pmatrix} 0 & 0 & 0 & i \\ 0 & 0 & -i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix}.$$

As a result, we have the following Theorem: There are isomorphisms

$$\text{Hom}(S^+, S^-) \cong \Lambda_{\mathbb{C}}^1, \quad \text{Hom}(S^-, S^+) \cong \Lambda_{\mathbb{C}}^1.$$

The elements in $C^2(V)$ correspond to the matrices

$$\begin{aligned} e_1 e_2 &= \begin{pmatrix} -i & 0 & 0 & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & i & 0 \\ 0 & 0 & 0 & -i \end{pmatrix} & e_3 e_4 &= \begin{pmatrix} -i & 0 & 0 & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & -i & 0 \\ 0 & 0 & 0 & i \end{pmatrix} \\ e_1 e_3 &= \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} & e_4 e_2 &= \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix} \\ e_1 e_4 &= \begin{pmatrix} 0 & i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & i \\ 0 & 0 & i & 0 \end{pmatrix} & e_2 e_3 &= \begin{pmatrix} 0 & i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & -i & 0 \end{pmatrix}. \end{aligned}$$

There is a representation

$$\text{Spin}(4) \longrightarrow \text{End}(\mathbb{C}^4), \quad \text{which splits into: } SU(2) \times SU(2).$$

4.13.2.3 (Anti) Self Duality

There is an operator on the exterior algebra which is similar to the chirality operator.

Let the *Hodge star operator*

$$\star : \Lambda V \longrightarrow \Lambda V,$$

with respect to Q , be given by the following relation:

$$e_{i_1} \wedge \cdots \wedge e_{i_k} \wedge \star(e_{i_1} \wedge \cdots \wedge e_{i_k}) = \epsilon_{i_1 \dots i_k} e_1 \wedge \cdots \wedge e_{\dim V},$$

where $\{e_i\}_{i=1, \dots, \dim V}$ is orthonormal with respect to Q and

$$\epsilon_{i_1 \dots i_k} = \text{sgn} \begin{pmatrix} 1 & 2 & \dots & \dim V \\ i_1 & \dots & i_k & i_{k+1} & \dots & i_{\dim V} \end{pmatrix}$$

with

$$(i_{k+1}, \dots, i_{\dim V}) = (1, 2, \dots, \widehat{i_1}, \dots, \widehat{i_k}, \dots, \dim V).$$

Now we consider the case when $\dim V = 4$.

For a four dimensional vector space V , the restriction of the square of the Hodge star operator satisfies:

$$\star^2|_{C^\pm(V)} = \pm 1_{\text{End}(C^\pm(V))}.$$

For a 4D vector space V , $\Lambda^2 V$ splits into ± 1 eigen-spaces of \star :

$$\Lambda_+^2 V = \left\{ \frac{(1 + \star)}{2} w \mid w \in \Lambda V \right\}$$

and

$$\Lambda_-^2 V = \left\{ \frac{(1 - \star)}{2} w \mid w \in \Lambda V \right\},$$

which are called the space of *self-dual* (SD) or *anti-self-dual* (ASD) 2-forms and are simplified as Λ_+ and Λ_- respectively.

The standard basis for Λ_+ and Λ_- are the *self-dual basis*:

$$\{w_1^+ = e_1 \wedge e_2 + e_3 \wedge e_4, \quad w_2^+ = e_1 \wedge e_3 + e_4 \wedge e_2, \quad w_3^+ = e_1 \wedge e_4 + e_2 \wedge e_3\},$$

and the *anti-self-dual basis*:

$$\{w_1^- = e_1 \wedge e_2 - e_3 \wedge e_4, \quad w_2^- = e_1 \wedge e_3 - e_4 \wedge e_2, \quad w_3^- = e_1 \wedge e_4 - e_2 \wedge e_3\},$$

respectively.

Notice that in the above corollary, the notation \mathbf{w}_i^\pm refers to elements of $C(V) \otimes \mathbb{C}$. If we consider \mathbf{w}_i^\pm as elements of $\Lambda_{\mathbb{C}}V = \Lambda V \otimes \mathbb{C}$, then we have

$$\mathbf{w}_1^+ = i(\bar{w}_1 \wedge w_1 + \bar{w}_2 \wedge w_2)$$

$$\mathbf{w}_2^+ = w_1 \wedge w_2 + \bar{w}_1 \wedge \bar{w}_2$$

$$\mathbf{w}_3^+ = i(w_1 \wedge w_2 - \bar{w}_1 \wedge \bar{w}_2)$$

$$\mathbf{w}_1^- = i(\bar{w}_1 \wedge w_1 - \bar{w}_2 \wedge w_2)$$

$$\mathbf{w}_2^- = \bar{w}_1 \wedge w_2 + w_1 \wedge \bar{w}_2$$

$$\mathbf{w}_3^- = i(\bar{w}_1 \wedge w_2 - w_1 \wedge \bar{w}_2).$$

The difference in the two interpretations of \mathbf{w}_1^+ arises from the fact that

$$\begin{aligned} w_i w_i &= \bar{w}_i \bar{w}_i = -1_{C(V)} \in C(V) \otimes \mathbb{C}, & \text{but} \\ w_i \wedge w_i &= \bar{w}_i \wedge \bar{w}_i = 0 \in \Lambda_{\mathbb{C}}V. \end{aligned}$$

By using the self-dual and anti-self dual basis, we can express our previous results again in a ‘better’ way:

The map $\tau: Spin(4) \longrightarrow SO(4)$ has the following images:

$$\tau(\exp(t \mathbf{w}_1^+)) = \begin{pmatrix} \cos 2t & -\sin 2t & 0 & 0 \\ \sin 2t & \cos 2t & 0 & 0 \\ 0 & 0 & \cos 2t & -\sin 2t \\ 0 & 0 & \sin 2t & \cos 2t \end{pmatrix}$$

$$\tau(\exp(t \mathbf{w}_2^+)) = \begin{pmatrix} \cos 2t & 0 & -\sin 2t & 0 \\ 0 & \cos 2t & 0 & \sin 2t \\ \sin 2t & 0 & \cos 2t & 0 \\ 0 & -\sin 2t & 0 & \cos 2t \end{pmatrix}$$

$$\tau(\exp(t \mathbf{w}_3^+)) = \begin{pmatrix} \cos 2t & 0 & 0 & -\sin 2t \\ 0 & \cos 2t & -\sin 2t & 0 \\ 0 & \sin 2t & \cos 2t & 0 \\ \sin 2t & 0 & 0 & \cos 2t \end{pmatrix}$$

$$\mathcal{T}(\exp(t w_1^-)) = \begin{pmatrix} \cos 2t & -\sin 2t & 0 & 0 \\ \sin 2t & \cos 2t & 0 & 0 \\ 0 & 0 & \cos 2t & \sin 2t \\ 0 & 0 & -\sin 2t & \cos 2t \end{pmatrix}$$

$$\mathcal{T}(\exp(t w_2^-)) = \begin{pmatrix} \cos 2t & 0 & -\sin 2t & 0 \\ 0 & \cos 2t & 0 & -\sin 2t \\ \sin 2t & 0 & \cos 2t & 0 \\ 0 & \sin 2t & 0 & \cos 2t \end{pmatrix}$$

$$\mathcal{T}(\exp(t w_3^-)) = \begin{pmatrix} \cos 2t & 0 & 0 & -\sin 2t \\ 0 & \cos 2t & \sin 2t & 0 \\ 0 & -\sin 2t & \cos 2t & 0 \\ \sin 2t & 0 & 0 & \cos 2t \end{pmatrix}.$$

The map $C^+(\mathbb{R}^4) \longrightarrow \mathbb{H} \oplus \mathbb{H}$ has the following images

$$\begin{aligned} 1 &\mapsto 1 \oplus 1, & \Gamma &= -e_1 e_2 e_3 e_4 \mapsto 1 \oplus -1 \\ w_1^+ &\mapsto -2i \oplus 0, & w_1^- &\mapsto 0 \oplus 2i \\ w_2^+ &\mapsto -2j \oplus 0, & w_2^- &\mapsto 0 \oplus -2j \\ w_3^+ &\mapsto 2k \oplus 0, & w_3^- &\mapsto 0 \oplus 2k, \end{aligned}$$

while the inverse $\mathbb{H} \oplus \mathbb{H} \longrightarrow \mathbb{C}^+(\mathbb{R}^4)$ has the following images

$$\begin{aligned} 1 \oplus 0 &\mapsto \frac{1 + \Gamma}{2}, & 0 \oplus 1 &\mapsto \frac{1 - \Gamma}{2} \\ i \oplus 0 &\mapsto \frac{-w_1^+}{2}, & 0 \oplus i &\mapsto \frac{w_1^-}{2} \\ j \oplus 0 &\mapsto \frac{-w_2^+}{2}, & 0 \oplus j &\mapsto \frac{-w_2^-}{2} \\ k \oplus 0 &\mapsto \frac{w_3^+}{2}, & 0 \oplus k &\mapsto \frac{w_3^-}{2}. \end{aligned}$$

So, the map $Spin(4) \longrightarrow SU(2) \times SU(2)$ has

$$\exp(t \Lambda_+) \longrightarrow SU(2) \times 1_{SU(2)}, \quad \exp(t \Lambda_-) \longrightarrow 1_{SU(2)} \times SU(2),$$

given by

$$\begin{aligned}
 \exp(t w_1^+) &\mapsto \begin{pmatrix} \exp(-i2t) & 0 \\ 0 & \exp(i2t) \end{pmatrix} \oplus \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\
 \exp(t w_2^+) &\mapsto \begin{pmatrix} \cos 2t & -\sin 2t \\ \sin 2t & \cos 2t \end{pmatrix} \oplus \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\
 \exp(t w_3^+) &\mapsto \begin{pmatrix} \cos 2t & \sin 2t i \\ \sin 2t i & \cos 2t \end{pmatrix} \oplus \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\
 \exp(t w_1^-) &\mapsto \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \oplus \begin{pmatrix} \exp(i2t) & 0 \\ 0 & \exp(-i2t) \end{pmatrix} \\
 \exp(t w_2^-) &\mapsto \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \oplus \begin{pmatrix} \cos 2t & -\sin 2t \\ \sin 2t & \cos 2t \end{pmatrix} \\
 \exp(t w_3^-) &\mapsto \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \oplus \begin{pmatrix} \cos 2t & \sin 2t i \\ \sin 2t i & \cos 2t \end{pmatrix}.
 \end{aligned}$$

The action of w_i^\pm in $C(V) \otimes \mathbb{C}$ on $S \cong \Lambda P$ is given by

$$\begin{aligned}
 w_1^+ &= 2 \begin{pmatrix} -i & 0 & 0 & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} & w_1^- &= 2 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & i & 0 \\ 0 & 0 & 0 & -i \end{pmatrix} \\
 w_2^+ &= 2 \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} & w_2^- &= 2 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \\
 w_3^+ &= 2 \begin{pmatrix} 0 & i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} & w_3^- &= 2 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & i \\ 0 & 0 & i & 0 \end{pmatrix}.
 \end{aligned}$$

Therefore, Λ_+ acts on S^+ while Λ_- acts on S^- .

Now consider the complexified algebra of self-dual and anti-self-dual two forms

$$\Lambda_{\pm\mathbb{C}} = \Lambda_{\pm} \otimes \mathbb{C}.$$

Usually, we use the following basis for $\Lambda_{+\mathbb{C}}$:

$$\begin{aligned}\bar{w}_1 \wedge w_1 + \bar{w}_2 \wedge w_2 &= -\mathrm{i} w_1^+ \\ w_1 \wedge w_2 &= \frac{w_2^+ - \mathrm{i} w_3^+}{2} \\ \bar{w}_1 \wedge \bar{w}_2 &= \frac{w_2^+ + \mathrm{i} w_3^+}{2}\end{aligned}$$

and the following basis for $\Lambda_{-\mathbb{C}}$:

$$\begin{aligned}\bar{w}_1 \wedge w_1 - \bar{w}_2 \wedge w_2 &= -\mathrm{i} w_1^- \\ \bar{w}_1 \wedge w_2 &= \frac{w_2^- - \mathrm{i} w_3^-}{2} \\ w_1 \wedge \bar{w}_2 &= \frac{w_2^- + \mathrm{i} w_3^-}{2}.\end{aligned}$$

Their quantization acts on S as

$$q(\bar{w}_1 \wedge w_1 + \bar{w}_2 \wedge w_2) = 2 \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$q(w_1 \wedge w_2) = 2 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$q(\bar{w}_1 \wedge \bar{w}_2) = 2 \begin{pmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$q(\bar{w}_1 \wedge w_1 - \bar{w}_2 \wedge w_2) = 2 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

$$q(\bar{w}_1 \wedge w_2) = 2 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$q(w_1 \wedge \bar{w}_2) = 2 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

As a result, we have the following Theorem: there are isomorphisms

$$\text{End}(S^+) \cong \Lambda_{\mathbb{C}}^0 \oplus \Lambda_{+\mathbb{C}}, \quad \text{End}(S^-) \cong \Lambda_{\mathbb{C}}^0 \oplus \Lambda_{-\mathbb{C}}.$$

4.13.2.4 Hermitian Structure on the Spinors

There is a canonical *Hermitian structure* on the space of positive spinors S^+ given by the *Hermitian inner product* $\langle \cdot, \cdot \rangle$, which takes the value

$$\langle s^+, t^+ \rangle = \bar{s}_1^+ t_1^+ + \bar{s}_2^+ t_2^+$$

on the spinors

$$s^+ = \begin{pmatrix} s_1^+ \\ s_2^+ \end{pmatrix} \quad \text{and} \quad t^+ = \begin{pmatrix} t_1^+ \\ t_2^+ \end{pmatrix} \in S^+.$$

The above Hermitian form is $Spin(4)$ -invariant.

The dual vector space S^{+*} consists of complex linear functionals

$$\phi : S^+ \longrightarrow \mathbb{C}.$$

S^{+*} is generated by the dual complex basis

$$\begin{aligned} & \{ 1_{C(V)}^*, (w_1 \wedge w_2)^* \}, \quad \text{which satisfies} \\ & 1_{C(V)}^*(1_{C(V)}) = 1, \quad 1_{C(V)}^*(w_1 \wedge w_2) = 0, \\ & (w_1 \wedge w_2)^*(1_{C(V)}) = 0, \quad (w_1 \wedge w_2)^*(w_1 \wedge w_2) = 1. \end{aligned}$$

There is a *Hermitian Riesz representation*

$$S^+ \xrightarrow{\cong} S^{+*},$$

with the following identification

$$\begin{pmatrix} s_1^+ \\ s_2^+ \end{pmatrix} \mapsto \left\langle \begin{pmatrix} s_1^+ \\ s_2^+ \end{pmatrix}, \cdot \right\rangle.$$

The Hermitian Riesz representation $S^+ \longrightarrow S^{+*}$ is given by

$$\begin{pmatrix} s_1^+ \\ s_2^+ \end{pmatrix} \mapsto \begin{pmatrix} \overline{s_1^+} \\ \overline{s_2^+} \end{pmatrix}^*,$$

which implies

$$\begin{pmatrix} s_1^+ \\ s_2^+ \end{pmatrix}^* = \left\langle \begin{pmatrix} \overline{s_1^+} \\ \overline{s_2^+} \end{pmatrix}, \cdot \right\rangle.$$

By using the Hermitian Riesz representation, we have

$$S^+ \otimes S^+ \cong \text{End}(S^+).$$

Also, by using the Hermitian Riesz representation, we have

$$S^+ \otimes S^+ \cong \Lambda_{\mathbb{C}}^0 \oplus \Lambda_{+\mathbb{C}}.$$

Now consider the negative spinors S^- . With respect to the standard basis $\{w_1, w_2\}$ we can define a similar *Spin(4)-invariant Hermitian inner product*, which takes the value

$$\langle s^-, t^- \rangle = \bar{s}_1^- t_1^- + \bar{s}_2^- t_2^-$$

on the spinors

$$s^- = \begin{pmatrix} s_1^- \\ s_2^- \end{pmatrix} \quad \text{and} \quad t^- = \begin{pmatrix} t_1^- \\ t_2^- \end{pmatrix} \in S^-.$$

We have a similar *Hermitian Riesz representation* on S^- :

$$S^- \xrightarrow{\cong} S^{-*},$$

with the following identification

$$\begin{pmatrix} s_1^- \\ s_2^- \end{pmatrix} \mapsto \left\langle \begin{pmatrix} s_1^- \\ s_2^- \end{pmatrix}, \cdot \right\rangle,$$

which, as in S^+ , satisfies

$$\begin{pmatrix} s_1^- \\ s_2^- \end{pmatrix}^* = \left\langle \begin{pmatrix} \bar{s}_1^- \\ \bar{s}_2^- \end{pmatrix}, \cdot \right\rangle.$$

As in S^+ , by using the Hermitian Riesz representation, we have:

$$S^- \otimes S^- \cong \text{End}(S^-),$$

$$S^- \otimes S^- \cong \Lambda_{\mathbb{C}}^0 \oplus \Lambda_{-\mathbb{C}},$$

$$S^+ \otimes S^- \xrightarrow{\cong} \text{Hom}(S^+, S^-),$$

$$S^+ \otimes S^- \cong \Lambda_{\mathbb{C}}^1.$$

Similarly, by interchanging S^+ and S^- , and by using the Hermitian Riesz representation, we have:

$$S^- \otimes S^+ \cong \text{Hom}(S^-, S^+),$$

$$S^- \otimes S^+ \cong \Lambda_{\mathbb{C}}^1.$$

4.13.2.5 Symplectic Structure on the Spinors

There is a canonical *symplectic structure* on the space of positive spinors S^+ given by the *symplectic form* $\{\cdot, \cdot\}$, which takes the value

$$\{s^+, t^+\} = s_1^+ t_2^+ - s_2^+ t_1^+$$

on the spinors

$$s^+ = \begin{pmatrix} s_1^+ \\ s_2^+ \end{pmatrix} \quad \text{and} \quad t^+ = \begin{pmatrix} t_1^+ \\ t_2^+ \end{pmatrix} \in S^+.$$

The above symplectic form is $Spin(4)$ -invariant.

There is a *symplectic Riesz representation*

$$S^+ \xrightarrow{\cong} S^{+*}$$

with the following identification

$$\begin{pmatrix} s_1^+ \\ s_2^+ \end{pmatrix} \mapsto \left\{ \begin{pmatrix} s_1^+ \\ s_2^+ \end{pmatrix}, \cdot \right\}.$$

The symplectic Riesz representation

$$S^+ \longrightarrow S^{+*}$$

is given by

$$\begin{pmatrix} s_1^+ \\ s_2^+ \end{pmatrix} \mapsto \begin{pmatrix} -s_2^+ \\ s_1^+ \end{pmatrix}^*.$$

That means

$$\begin{pmatrix} s_1^+ \\ s_2^+ \end{pmatrix}^* = \left\{ \begin{pmatrix} s_2^+ \\ -s_1^+ \end{pmatrix}, \cdot \right\}.$$

Just like the Hermitian case, we also have the following.

By using the symplectic Riesz representation, we have

$$S^+ \otimes S^+ \cong \text{End}(S^+),$$

$$S^+ \otimes S^+ \cong \Lambda_{\mathbb{C}}^0 \oplus \Lambda_{+\mathbb{C}}.$$

Now consider the negative spinors S^- . With respect to the standard basis $\{w_1, w_2\}$ we can define a similar *Spin(4)-invariant symplectic form* which takes the value

$$\{s^-, t^-\} = s_1^- t_2^- - s_2^- t_1^-$$

on the spinors

$$s^- = \begin{pmatrix} s_1^- \\ s_2^- \end{pmatrix} \quad \text{and} \quad t^- = \begin{pmatrix} t_1^- \\ t_2^- \end{pmatrix} \in S^-.$$

We have a similar *symplectic Riesz representation* on S^- :

$$S^- \xrightarrow{\cong} S^{-*}$$

with the following identification

$$\begin{pmatrix} s_1^- \\ s_2^- \end{pmatrix} \mapsto \left\{ \begin{pmatrix} s_1^- \\ s_2^- \end{pmatrix}, \cdot \right\},$$

which, as in S^+ , satisfies

$$\begin{pmatrix} s_1^- \\ s_2^- \end{pmatrix}^* = \left\{ \begin{pmatrix} s_2^- \\ -s_1^- \end{pmatrix}, \cdot \right\}.$$

As in S^+ , by using the symplectic Riesz representation, we have:

$$S^- \otimes S^- \cong \text{End}(S^-),$$

$$S^- \otimes S^- \cong \Lambda_{\mathbb{C}}^0 \oplus \Lambda_{-\mathbb{C}},$$

$$S^+ \otimes S^- \xrightarrow{\cong} \text{Hom}(S^+, S^-),$$

$$S^+ \otimes S^- \cong \Lambda_{\mathbb{C}}^1.$$

Similarly, by interchanging S^+ and S^- , and by using the symplectic Riesz representation, we have

$$S^- \otimes S^+ \cong \text{Hom}(S^-, S^+),$$

$$S^- \otimes S^+ \cong \Lambda_{\mathbb{C}}^1.$$

4.13.3 Penrose Twistor Calculus

Recall that the twistor theory, originally developed by Roger Penrose in 1967, is the mathematical theory which maps the geometric objects of the 4D Minkowski space-time into the geometric objects in the 4D complex space with the metric signature $(2, 2)$. The coordinates in such a space are called *twistors*.

The twistor approach appears to be especially natural for solving the equations of motion of massless fields of arbitrary spin.

Recently, Ed Witten used twistor theory to understand certain Yang–Mills amplitudes, by relating them to a certain string theory, the topological B model, embedded in twistor space. This field has come to be known as twistor string theory.

4.13.3.1 Penrose Index Formalism

Except where otherwise indicated we use Penrose’s abstract index notation [Penrose and Rindler (1984)] which allows for easy explicit calculations without involving a choice of basis. Thus we may write, v^A or v^B for a section of the unprimed fundamental spinor bundle \mathcal{E}^A . Similarly $w_{A'}$ could denote a section of the primed fundamental spinor bundle $\mathcal{E}_{A'}$. We write \mathcal{E}_A for the dual bundle to \mathcal{E}^A and $\mathcal{E}^{A'}$ for the dual to $\mathcal{E}_{A'}$. The tensor products of these bundles yield the general spinor objects such as $\mathcal{E}_{AB} := \mathcal{E}_A \otimes \mathcal{E}_B$, $\mathcal{E}_{A'B'}^{ABC'}$ and so forth. The tensorial indices are also abstract indices. Recall that $\mathcal{E}^a = \mathcal{E}_{A'}^A$ is the tangent bundle, so $\mathcal{E}_a = \mathcal{E}_A^{A'}$ is the cotangent bundle and we may use the terms ‘spinor’ or ‘section of a spinor bundle’ to describe tensor fields [Gover and Slovak (1999)].

A spinor object on which some indices have been contracted will be termed a *spinor contraction* (of the underlying spinor). For example, $v_{BC'DE}^{ABC'}$ is a contraction of $v_{DD'E'F}^{ABC'}$. In many cases the underlying spinor of interest is a tensor product of lower valence spinors. For example, $v^{AB}w_B^{C'}u_{ACD}$ is a contraction of $v^{AB}w_C^{C'}u_{DEF}$. The same conventions are used for the tensor indices and the twistor indices; the latter are to introduced below. Standard notation is also used for the symmetrizations and antisymmetrizations over some indices.

Weights and scales

We define line bundles of densities or *weighted functions* as follows [Gover and Slovak (1999)]. The weight -1 *line bundle* $\mathcal{E}[-1]$ over M is identified

with $\mathcal{E}[\overbrace{A'B'\cdots C'}^p]$. Then, for integral w , the weight w line bundle $\mathcal{E}[w]$ is defined to be $(\mathcal{E}[-1])^{-w}$. In the case of AG-geometries corresponding to the real-split form $SL(p+q, \mathbb{R})$ we can (locally) extend this definition to weights $w \in \mathbb{R}$, by locally selecting a ray fibre subbundle of $\mathcal{E}[-1]$. Calling this say $\mathcal{E}_+[-1]$, we can then define the *ray bundles* $\mathcal{E}_+[w] := (\mathcal{E}_+[-1])^{-w}$. Finally these may be canonically extended to line bundles in the obvious way. In any case we write $\mathcal{E}^{A'}[w]$ for $\mathcal{E}^{A'} \otimes \mathcal{E}[w]$ and so on, whenever defined. In view of the defining isomorphism

$$h : \wedge^q \mathcal{E}^A \xrightarrow{\cong} \wedge^p \mathcal{E}_{A'}, \quad \text{we also have} \quad (4.165)$$

$$\mathcal{E}[-1] \cong \mathcal{E}[\underbrace{AB\cdots C}_q], \quad \mathcal{E}[1] \cong \mathcal{E}[\underbrace{AB\cdots C}_q] \cong \mathcal{E}[\underbrace{A'B'\cdots C'}_p].$$

We write $\epsilon^{A'B'\cdots C'}$ for the *tautological section* of $\mathcal{E}^{[A'B'\cdots C']}[1]$ giving the mapping $\mathcal{E}[-1] \xrightarrow{\cong} \wedge^p \mathcal{E}^{A'}$ by

$$f \mapsto f \epsilon^{A'B'\cdots C'}, \quad (4.166)$$

and $\epsilon_{D\cdots E}$ for similar object giving $\mathcal{E}[-1] \xrightarrow{\cong} \wedge^q \mathcal{E}_A$. A *scale* for the AG-structure is a nowhere vanishing section ξ of $\mathcal{E}[1]$. Note that such a choice is equivalent to a choice of spinor ‘volume’ form

$$\epsilon_\xi^{A'\cdots C'} := \xi^{-1} \epsilon^{A'\cdots C'}, \quad \text{or to a choice of form,} \quad \epsilon_{D\cdots E}^\xi := \xi^{-1} \epsilon_{D\cdots E}.$$

Distinguished connections

A connection ∇_a on M belongs to the given AG-structure (this really means ∇_a comes from a principal connection on the bundle \mathcal{G}_0 described below) iff it satisfies two conditions [Gover and Slovak (1999)]:

- ∇_a is the tensor product of linear connections (both of which we shall also denote ∇_a) on the spinor bundles \mathcal{E}^A and $\mathcal{E}_{A'}$, and
- The defining isomorphism h in (4.165) is covariantly constant, i.e., $\nabla_a h = 0$.

Our conventions for the torsion $T_{ab}{}^c$ and curvature $R_{ab}{}^c{}_d$ of a connection ∇_a on the tangent bundle TM are determined by the following equation,

$$2\nabla_{[a}\nabla_{b]}v^c = T_{ab}{}^d\nabla_d v^c + R_{ab}{}^c{}_d v^d.$$

Since $T_{ab}{}^c$ is skew on its lower indices, $T_{ab}{}^c = T_{[ab]}{}^c$, it can be written as a sum of two terms

$$T_{ab}{}^c = F_{ab}{}^c + \tilde{F}_{ab}{}^c, \quad \text{where} \\ F_{ab}{}^c = : F_{ABC'}^{A'B'C} = F_{(AB)C'}^{[A'B']C}, \quad \tilde{F}_{ab}{}^c = : \tilde{F}_{ABC'}^{A'B'C} = F_{[AB]C'}^{(A'B')C}.$$

The Cartan bundle \mathcal{G} over the manifold M has the quotient \mathcal{G}_0 , a principal fibre bundle with structure group G_0 . By the general theory, each G_0 -equivariant section $\sigma : \mathcal{G}_0 \rightarrow \mathcal{G}$ of the quotient projection defines the distinguished principal connection on \mathcal{G}_0 , the pullback of the \mathfrak{g}_0 -part of ω . The whole class of these connections consists precisely of connections on \mathcal{G}_0 with the unique torsion taking values in the kernel of ∂^* . A straightforward computation shows that the latter condition is equivalent to the condition that both \tilde{F} and F be completely trace-free. Each principal connection on \mathcal{G}_0 induces the induced connection on the bundle $\mathcal{E}[1] \setminus \{0\}$ which is associated to \mathcal{G}_0 and, moreover, the resulting correspondence between the sections σ and the latter connections is bijective. In particular, each section ξ of the bundle $\mathcal{E}[1] \setminus \{0\}$ defines uniquely a reduction σ , such that the corresponding distinguished connection leaves ξ horizontal.

Therefore, given a scale ξ on an AG-structure there are unique connections on \mathcal{E}^A and $\mathcal{E}_{A'}$ such that $F_{ABC'}^{A'B'C}$ and $\tilde{F}_{ABC'}^{A'B'C}$ are totally trace-free, the induced covariant derivative preserves the isomorphism h of (4.165), and $\nabla_a \xi = 0$. The torsion components $F_{ab}{}^c$ and $\tilde{F}_{ab}{}^c$ of the induced connection on TM are invariants of the so-called *AG-structures* [Bailey and Eastwood (1991)].

Note that in the special case of the four-dimensional conformal geometries, there is always a connection with vanishing torsion on \mathcal{G}_0 and so both F and \tilde{F} are zero. The scales correspond to a choice of metric from the conformal class while the general distinguished connections (corresponding to the reduction parameter σ being not necessarily exact) are just the Weyl geometries.

We may write ∇_a^ξ to indicate a connection as determined by the Theorem, although mostly we omit the ξ . Thus we might write $\nabla_a^{\hat{\xi}}$ or simply $\hat{\nabla}_a$ to indicate the connection corresponding to a scale $\hat{\xi}$ and similar conventions will be used for other operators and tensors that depend on ξ .

In what follows, for the purpose of explicit calculations, we shall often choose a scale and work with the corresponding connections. Objects are then well defined, or *invariant*, (on the AG-structure) if they are independent of the choice of scale. Note that if we change the scale according to

$\xi \mapsto \hat{\xi} = \Omega^{-1}\xi$, where Ω is a smooth non-vanishing function, then the connection transforms as follows [Gover and Slovak (1999)]:

$$\begin{aligned}\mathcal{E}^A : \hat{\nabla}_A^{A'} u^C &= \nabla_A^{A'} u^C + \delta_A^C \Upsilon_B^{A'} u^B \\ \mathcal{E}_{A'} : \hat{\nabla}_A^{A'} u_{C'} &= \nabla_A^{A'} u_{C'} + \delta_{C'}^{A'} \Upsilon_A^{B'} u_{B'} \\ \mathcal{E}_B : \hat{\nabla}_A^{A'} v_B &= \nabla_A^{A'} v_B - \Upsilon_B^{A'} v_A \\ \mathcal{E}^{B'} : \hat{\nabla}_A^{A'} v^{B'} &= \nabla_A^{A'} v^{B'} - \Upsilon_A^{B'} v^{A'}\end{aligned}\quad (4.167)$$

where $\Upsilon_a := \Omega^{-1} \nabla_a \Omega$. Consequently

$$\hat{\nabla}_a f = \nabla_a f + w \Upsilon_a f \quad \text{if } f \in \mathcal{E}[w]. \quad (4.168)$$

Given a choice of scale ξ , we write R_{abD}^C (or $R_{abD}^{(\xi)C}$ to emphasise the choice of scale) for the curvature of ∇_a on \mathcal{E}^A and $R_{abD'}^{C'}$ for the curvature of ∇_a on $\mathcal{E}_{A'}$, that is

$$(2\nabla_{[a} \nabla_{b]} - T_{ab}{}^e \nabla_e) v^C = R_{abD}^C v^D, \quad (2\nabla_{[a} \nabla_{b]} - T_{ab}{}^e \nabla_e) w_{D'} = -R_{abD'}^{C'} w_{C'}.$$

Then the curvature of the induced linear connection on TM is

$$R_{ab}{}^c{}_d = R_{abD}^{C'} \delta_D^C + R_{abD}^C \delta_{D'}^{C'}.$$

Note that since ∇_a preserves the volume forms $\epsilon_{A' \dots C'}$ and $\epsilon_{D \dots E}$ it follows that R_{abD}^C and $R_{abD'}^{C'}$ are trace-free on the spinor indices displayed. Thus the equations

$$R_{abD}^C = U_{abD}^C - \delta_B^C P_{AD}^{A'B'} + \delta_A^C P_{BD}^{B'A'}, \quad R_{abD'}^{C'} = U_{abD'}^{C'} + \delta_{D'}^{B'} P_{AB}^{A'C'} - \delta_{D'}^{A'} P_{BA}^{B'C'}$$

determine the objects U_{abD}^C , $U_{abD'}^{C'}$ and the *Rho-tensor*, P_{ab} , if we require that $U_{ACD}^{A'B'C} = 0 = U_{ABD'}^{A'B'C'}$. In this notation we have,

$$\begin{aligned}R_{ab}{}^c{}_d &= U_{ab}{}^c{}_d + \delta_{C'}^{D'} \delta_A^C P_{BD}^{B'A'} - \delta_{C'}^{D'} \delta_B^C P_{AD}^{A'B'} - \delta_D^C \delta_{C'}^{A'} P_{BA}^{B'D'} + \delta_D^C \delta_{C'}^{B'} P_{AB}^{A'D'} \\ \text{where } U_{ab}{}^c{}_d &= U_{abD}^C \delta_{C'}^{D'} + U_{abD'}^{C'} \delta_D^C.\end{aligned}\quad (4.170)$$

In the case of $p = 2 = q$ this agrees with the usual decomposition of the curvature of the Levi-Civita connection into the conformally invariant (and trace-free) Weyl tensor part and the remaining part given by the Rho-tensor (see e.g., [Bailey *et al.* (1994)]). Note that U 's are 2-forms valued in \mathfrak{g}_0 coming from the curvature of the canonical Cartan connection and so they are in the kernel of ∂^* . This is the source of the condition on the trace, but they are not trace-free in general:

$$U_{abC}^C = -U_{abC'}^{C'} = 2P_{[ab]}. \quad (4.171)$$

On the other hand, it follows from the Bianchi identity,

$$R_{[abc]}^d + \nabla_{[a} T_{bc]}^d + T_{[ab}^e T_{c]e}^d = 0, \quad \text{that} \\ 2(p+q)P_{[ab]} = -\nabla_c T_{ab}^c.$$

The Rho-tensor P_{ab} has the transformation equation

$$\hat{P}_{AB}^{A'B'} = P_{AB}^{A'B'} - \nabla_A^{A'} \Upsilon_B^{B'} + \Upsilon_A^{B'} \Upsilon_B^{A'}. \quad (4.172)$$

We are most interested in the special case $p = 2$. Then the whole component F_{ab}^c is irreducible and so it vanishes by our condition on the trace, while the other component \tilde{F}_{ab}^c of the torsion, together with the trace-free part of $U_{(ABC)}^{[A'B']D}$ are the only local invariants of the structures. In all other cases $2 < p \leq q$, the two components of the torsion are the only invariants.

The totally symmetrized covariant derivatives of the Rho-tensors play a special role. We use the notation

$$S_{a \cdots b} := \underbrace{\nabla_{(a} \nabla_b \cdots \nabla_d P_{ef)}}_s \quad \text{for } s = 2, 3, \dots.$$

Twistors

Via the Cartan bundle \mathcal{G} over M any P -module V gives rise to a *natural bundle* (or induced bundle) \mathcal{V} . Sections of \mathcal{V} are identified with functions $f : \mathcal{G} \rightarrow V$ such that $f(x.p) = \rho(p^{-1})f(x)$, where $x \mapsto x.p$ gives the action of $p \in P$ on $x \in \mathcal{G}$ while ρ is the action defining the P -module structure.

Recall also that the Cartan bundle is equipped with a canonical connection, the so called normal Cartan connection ω . In view of this it is in our interests to work, where possible, with natural bundles \mathcal{V} induced from V where this is not merely a P -module but in fact a G -module. Then the Cartan connection induces an invariant linear connection on \mathcal{V} . Let us write V^α for the module corresponding to the standard representation of G on \mathbb{R}^{p+q} and write V_α for the dual module. The index α is another Penrose-type abstract index and we write \mathcal{E}^α and \mathcal{E}_α for the respective bundles induced by these G -modules. All finite dimensional G -modules are submodules in tensor products of the fundamental representations V^α and V_α . Thus the bundles \mathcal{E}^α and \mathcal{E}_α play a special role and we term these (*local*) *twistor bundles* [Bailey and Eastwood (1991); Penrose and Rindler (1986)]. In fact in line with the use of the word “tensor” we also describe any explicit subbundle of a tensor product of

these bundles as a twistor bundle and sections of such bundles as local twistors. In particular observe that there is a canonical completely skew local-twistor $(p+q)$ -form $h_{\alpha\beta\cdots\gamma}$ on \mathcal{E}^α which is equivalent to the isomorphism (4.165). We write $h^{\alpha\beta\cdots\gamma}$ for the dual completely skew twistor satisfying $h^{\alpha\beta\cdots\gamma}h_{\alpha\beta\cdots\gamma} = (p+q)!$.

All finite dimensional P -modules enjoy filtrations which split completely as G_0 -modules. V^α and V_α , give the simplest cases and, as P -modules, admit filtrations

$$V^\alpha = V^A + V^{A'}, \quad V_\alpha = V_{A'} + V_A.$$

(Our notational convention is that the ‘right ends’ in the formal sums are submodules while the ‘left ends’ are quotients.) These determine filtrations of the twistor bundles

$$\mathcal{E}^\alpha = \mathcal{E}^A + \mathcal{E}^{A'}, \quad \mathcal{E}_\alpha = \mathcal{E}_{A'} + \mathcal{E}_A.$$

We write $X_{A'}^\alpha$ for the canonical section of $\mathcal{E}_{A'}^\alpha$ which gives the injecting morphism $\mathcal{E}^{A'} \rightarrow \mathcal{E}^\alpha$ via

$$v^{A'} \mapsto X_{A'}^\alpha v^{A'}. \quad (4.173)$$

Similarly Y_α^A describes the injection of \mathcal{E}_A into dual twistors,

$$\mathcal{E}_A \ni u_A \mapsto Y_\alpha^A u_A \in \mathcal{E}_\alpha. \quad (4.174)$$

It follows from standard representation theory that a choice of splitting of the exact sequence,

$$0 \rightarrow V^{A'} \rightarrow V^\alpha \rightarrow V^A \rightarrow 0$$

is equivalent to the choice of subgroup of P which is isomorphic to G_0 . It follows immediately that a choice of splitting of the twistor bundle \mathcal{E}^α is equivalent to a reduction from \mathcal{G} to \mathcal{G}_0 . Such a splitting is a G_0 -equivariant homomorphism $\xi : \mathcal{E}^\alpha \rightarrow \mathcal{E}^{A'}$. We can regard ξ here as a section of $\mathcal{E}_\alpha \otimes \mathcal{E}^{A'} = \mathcal{E}_\alpha^{A'}$ and then in our index notation the homomorphism is determined by $v^\alpha \mapsto \xi_\alpha^{A'} v^\alpha$, for any section v^α of \mathcal{E}^α . The composition of ξ with the monomorphism $\mathcal{E}^{A'} \rightarrow \mathcal{E}^\alpha$ must be the identity so we have,

$$\xi_\beta^{A'} X_{B'}^\beta = \delta_{B'}^{A'}.$$

A splitting $\xi_\alpha^{A'}$ of \mathcal{E}^α determines a dual splitting λ_A^α of \mathcal{E}_α , $\lambda_A^\alpha : \mathcal{E}_\alpha \rightarrow \mathcal{E}_A$. Given such splittings we have $\mathcal{E}_\alpha = \mathcal{E}^A \oplus \mathcal{E}^{A'}$ and $\mathcal{E}_\alpha = \mathcal{E}_{A'} \oplus \mathcal{E}_A$, so we

may write sections of these bundles as a “matrices” such as

$$[u^\alpha]_\xi = \begin{pmatrix} u^A \\ u^{A'} \end{pmatrix} \in [\Gamma \mathcal{E}^\alpha]_\xi \quad [v_\alpha]_\xi = (v_A :: v_{A'}) \in [\Gamma \mathcal{E}_\alpha]_\xi.$$

We will always work with splittings determined by a choice of scale $\xi \in \mathcal{E}[1]$, as discussed earlier. If u^α and v_α , as displayed, are expressed by such a scale then the change of scale $\xi \mapsto \hat{\xi} = \Omega^{-1}\xi$ yields a transformation of these splittings. For example $[u^\alpha] \mapsto [u^\alpha]_{\hat{\xi}}$ where

$$[u^\alpha]_{\hat{\xi}} = \begin{pmatrix} \hat{u}^A \\ \hat{u}^{A'} \end{pmatrix} = \begin{pmatrix} u^A \\ u^{A'} - \Upsilon_B^{A'} u^B \end{pmatrix}.$$

With this understood we henceforth drop the notation $[\cdot]_\xi$ and simply write, for example, $v_\alpha \mapsto \hat{v}_\alpha$ where

$$\hat{v}_\alpha = (\hat{v}_A :: \hat{v}_{A'}) = (v_A + \Upsilon_A^{B'} v_{B'} :: v_{A'}),$$

for the corresponding transformation of v^α . In particular, the objects $\xi_\alpha^{B'}$, λ_A^β are not invariant and

$$\hat{\xi}_\alpha^{B'} = \xi_\alpha^{B'} - Y_\alpha^{A'} \Upsilon_A^{B'}, \quad \hat{\lambda}_A^\beta = \lambda_A^\beta + X_{B'}^\beta \Upsilon_A^{B'}.$$

However, note that, in the splittings they determine, $\xi_\alpha^{B'}$ and λ_A^β are given

$$\xi_\alpha^{B'} = \begin{pmatrix} 0 \\ \delta_{A'}^{B'} \end{pmatrix}, \quad \lambda_A^\beta = \begin{pmatrix} \delta_A^\beta \\ 0 \end{pmatrix}.$$

In any such splitting the invariant objects $X_{B'}^\alpha$ and Y_β^A are given by

$$X_{B'}^\alpha = \begin{pmatrix} 0 \\ \delta_{B'}^{A'} \end{pmatrix}, \quad Y_\beta^A = \begin{pmatrix} \delta_B^A \\ 0 \end{pmatrix}.$$

The first four identities of the following display are immediate, while the final two items are useful definitions:

$$\begin{aligned} Y_\beta^A X_{A'}^\beta &= 0, & \xi_\beta^{A'} \lambda_A^\beta &= 0, \\ Y_\beta^A \lambda_B^\beta &= \delta_B^A, & \xi_\beta^{A'} X_{B'}^\beta &= \delta_{B'}^{A'}, \\ Y_\beta^A \lambda_A^\gamma &=: \lambda_\beta^\gamma, & \xi_\beta^{A'} X_{A'}^\gamma &=: \xi_\beta^\gamma. \end{aligned} \tag{4.175}$$

We shall mostly deal with *weighted twistors*, i.e., tensor products of the form $\mathcal{E}_{\gamma \dots \delta}^{\alpha \dots \beta}[w] = \mathcal{E}_{\gamma \dots \delta}^{\alpha \dots \beta} \otimes \mathcal{E}[w]$. All the above algebraic machinery works for the weighted twistors. In fact we shall often omit the word ‘weighted’ even though, of course, these bundles do not come from G -modules for $w \neq 0$.

Finally, we observe that via this machinery any spinorial quantity may be identified with a (weighted) twistor. For example, valence 1 spinors in $\mathcal{E}^{A'}[w_1]$ or $\mathcal{E}_A[w_2]$ may be dealt with via (4.173) or (4.174) respectively. This determines an identification for tensor powers by treating each factor in this way. This does all cases since, via (4.166),

$$\mathcal{E}^A \cong \mathcal{E}_{\underbrace{B \dots D}_{q-1}}[1], \quad \mathcal{E}_{A'} \cong \mathcal{E}^{\overbrace{B' \dots C'}^{p-1}}[-1].$$

Now, any irreducible representation of G_0 is given as a tensor product of two irreducible components in tensor products of the fundamental spinors (viewed as representations of the special linear groups, adjusted by a weight). Applying the corresponding Young symmetrizers [Penrose and Rindler (1984); Fulton and Harris (1991)] to the tensor products of \mathcal{E}_α and \mathcal{E}^β , we get the explicit realization of each irreducible spinor bundle as the subbundle of the (weighted) twistor bundle which is isomorphic to the injecting part of the twistor bundle. Thus a section of a weighted irreducible spinor bundle \mathcal{V} may be identified with a twistor object which is zero in all its composition factors except the first. So, in fact, this non-zero factor is also the projecting part of the twistor. We write $\tilde{\mathcal{V}}$ for this twistor (sub-)bundle satisfying $\mathcal{V} \cong \tilde{\mathcal{V}}$. Altogether, we have established the following result [Gover and Slovak (1999)]: Any irreducible spinor object v can be identified with the twistor \tilde{v} which has the spinor as its projecting part. This identification is provided in a canonical algebraic way.

4.13.3.2 Twistor Calculus

Given a choice of scale ξ , a *twistor connection* ∇_a on \mathcal{E}^α and \mathcal{E}_α is given by [Penrose and MacCallum (1972); Dighton (1974); Bailey and Eastwood (1991)]

$$\nabla_A^{P'} \begin{pmatrix} v^B \\ v^{B'} \end{pmatrix} = \begin{pmatrix} \nabla_A^{P'} v^B + \delta_A^B v^{P'} \\ \nabla_A^{P'} v^{B'} - P_{AB}^{P'} v^B \end{pmatrix} \quad (4.176)$$

and

$$\nabla_A^{P'}(u_B :: u_{B'}) = (\nabla_A^{P'} u_B + P_{AB}^{P'} u_{B'} :: \nabla_A^{P'} u_{B'} - \delta_{B'}^A u_B). \quad (4.177)$$

Notice that whereas on the left hand side ∇ indicates the twistor connection, on the right hand side the symbol ∇ indicates the usual spinor connection

determined by the choice of scale. Although we have fixed a choice of scale to present explicit formulae for these connections, it is easily verified directly using the formulae (4.167) that the twistor connections are in fact independent of the choice of scale.

A straightforward calculation gives

$$\begin{aligned} ([\nabla_a, \nabla_b] - T_{ab}{}^d \nabla_d) \begin{pmatrix} v^C \\ v^{C'} \end{pmatrix} &= \\ &= \begin{pmatrix} U_{abD}{}^C v^D - T_{abD'}{}^{C'} v^{D'} \\ -2\nabla_{[a} P_{b]D}{}^{C'} v^D + T_{abE'}{}^{D'} P_{ED}{}^{C'} v^D + U_{abD'}{}^{C'} v^{D'} \end{pmatrix}. \end{aligned}$$

Thus the curvature of the twistor connection is given, in this scale, by

$$W_{ab\delta}{}^\gamma = \begin{pmatrix} U_{abD}{}^C & -T_{abD'}{}^{C'} \\ -2Q_{abD}{}^{C'} & U_{abD'}{}^{C'} \end{pmatrix}, \quad \text{where} \quad Q_{abc} := \nabla_{[a} P_{b]c} - \frac{1}{2} T_{ab}{}^e P_{ec}. \quad (4.178)$$

Note that since the twistor connection is invariant it follows that this *twistor curvature* $W_{ab\delta}{}^\gamma$ is invariant. In fact, viewed as a \mathfrak{g} -valued 2-form on the Cartan bundle \mathcal{G} , this is just the curvature of the normal Cartan connection. In particular, we know that the structures are torsion-free (in the sense of the Cartan connection) if and only if the torsion part $T_{ab}{}^c$ vanishes and they are locally flat if and only if the whole $W_{ab\delta}{}^\gamma$ vanishes.

The D -operators

If $f \in \mathcal{E}[w]$ then it follows easily from (4.168) that the *spinor-twistor object*

$$D_\beta^{A'} f := (\nabla_B^{A'} f \ w \delta_B^{A'} f)$$

is invariant. We may regard this as an injecting part of the invariant twistor object $D_\beta^\alpha f := X_{A'}^\alpha D_\beta^{A'} f$. By regarding, in this formula for D_β^α , ∇ to be the coupled twistor-spinor connection it is easily verified that the operator D_β^α is well defined and invariant on sections of the weighted twistor bundles $\mathcal{E}_{\alpha \dots \gamma}^{\rho \dots \mu}[w]$. The invariant operators $D_\beta^\alpha : \mathcal{E}_{\delta \dots \gamma}^{\rho \dots \mu}[w] \rightarrow \mathcal{E}_{\beta \delta \dots \gamma}^{\alpha \rho \dots \mu}[w]$ are called the *twistor- D operators*.

For many calculations, where a choice of scale is made, it is useful to allow D_β^α to operate on spinors and their tensor products, although in this case the result is not independent of the scale. For example, if $v_C \in \mathcal{E}_C[w]$ then

$$D_\beta^{A'} v_C := (\nabla_B^{A'} v_C \ w \delta_B^{A'} v_C).$$

Since the operator D_β^α and its concatenations will have an important role in the following discussions we develop notation for their target spaces. First let \mathcal{F}^ρ be defined as follows,

$$\mathcal{F}^\rho := \text{Ker}(Y_\rho^A :: \mathcal{E}^\rho \rightarrow \mathcal{E}^A).$$

Then we write

$$\mathcal{F}_{\alpha \dots \beta}^{\rho \dots \sigma} := \mathcal{F}^\rho \otimes \dots \otimes \mathcal{F}^\sigma \otimes \mathcal{E}_\alpha \otimes \dots \otimes \mathcal{E}_\beta,$$

and $\mathcal{F}_{\alpha \dots \beta}^{\rho \dots \sigma}[w] = \mathcal{F}_{\alpha \dots \beta}^{\rho \dots \sigma} \otimes \mathcal{E}[w]$. Finally let

$$\underbrace{\mathcal{S}_{\alpha \dots \beta}^{\rho \dots \sigma}}_k[w] := (\odot^k \mathcal{F}_\alpha^\rho) \otimes \mathcal{E}[w].$$

Note that sections of $\mathcal{F}_\rho^\alpha (= \mathcal{S}_\rho^\alpha)$ are not generally trace-free, but that \mathcal{F}_ρ^α is in a complement to the trace-part of \mathcal{E}_ρ^α .

Now if $f \in \mathcal{E}[w]$ then $D_\alpha^\rho f \in \mathcal{F}_\alpha^\rho[w]$. Similarly observe that if $v^\sigma \in \mathcal{F}^\sigma$ then

$$D_\alpha^\rho v^\sigma - \delta_\alpha^\sigma v^\rho$$

is in $\mathcal{F}_\alpha^{\rho\sigma}$. Thus

$$D_{\alpha\beta}^{\rho\sigma} := \frac{1}{2} (D_\alpha^\rho D_\beta^\sigma + D_\beta^\sigma D_\alpha^\rho - \delta_\alpha^\sigma D_\beta^\rho - \delta_\beta^\rho D_\alpha^\sigma)$$

gives an invariant operator

$$D_{\alpha\beta}^{\rho\sigma} : \mathcal{E}_{\gamma \dots \delta}^{\mu \dots \nu}[w] \rightarrow \mathcal{S}_{\alpha\beta}^{\rho\sigma} \otimes \mathcal{E}_{\gamma \dots \delta}^{\mu \dots \nu}[w].$$

Similarly we define $D_{\alpha\beta\gamma}^{\rho\sigma\mu}$ by

$$D_{\alpha\beta\gamma}^{\rho\sigma\mu} := \frac{1}{3} ((D_\alpha^\rho D_{\beta\gamma}^{\sigma\mu} + D_\beta^\sigma D_{\alpha\gamma}^{\rho\mu} + D_\gamma^\mu D_{\alpha\beta}^{\rho\sigma} - \delta_\alpha^\rho D_{\beta\gamma}^{\sigma\mu} - \delta_\beta^\sigma D_{\alpha\gamma}^{\rho\mu} - \delta_\gamma^\mu D_{\alpha\beta}^{\rho\sigma})$$

and so on for $D_{\rho \dots \nu}^{\alpha \dots \delta}$. Notice that the construction of these is designed in such a way that the resulting operators are annihilated if composed (contracted) with Y_ν^B on any index.

The Splitting Machinery

In terms of the algebraic projectors and embeddings introduced in the last

section, the twistor-D operator is given by

$$D_\alpha^\rho f = X_{R'}^\rho Y_\alpha^A \nabla_A^{R'} f + w \xi_\alpha^\rho f, \quad (4.179)$$

where f is any weighted twistor-spinor object. Using this and the expressions (4.176), (4.177) for the twistor connection, the following identities are easily established:

$$\begin{aligned} D_\alpha^\rho X_{C'}^\beta &= X_{C'}^\rho \lambda_K^\beta Y_\alpha^K & D_\alpha^\rho Y_\beta^C &= -Y_\alpha^C X_{K'}^\rho \xi_\beta^{K'} \\ D_\alpha^\rho \xi_\beta^{S'} &= P_{\alpha\beta}^{\rho S'} & D_\alpha^\rho \lambda_B^\sigma &= -P_{\alpha B}^{\rho\sigma} \\ X_{B'}^\alpha D_\alpha^\gamma f &= w X_{B'}^\gamma f & Y_\gamma^B D_\alpha^\gamma f &= 0 \\ \xi_\gamma^{B'} D_\alpha^\gamma f &= D_\alpha^{B'} f & \lambda_B^\alpha D_\alpha^{B'} f &= \nabla_B^{B'} f, \end{aligned} \quad (4.180)$$

where, again, f is any weighted twistor-spinor and we write

$$P_{\alpha\beta}^{\rho\sigma} := P_{AB}^{R'S'} X_{R'}^\rho X_S^\sigma Y_\alpha^A Y_\beta^B, \quad P_{\alpha\beta}^{\rho S'} := P_{AB}^{R'S'} X_{R'}^\rho Y_\alpha^A Y_\beta^B, \quad P_{\alpha B}^{\rho\sigma} := P_{AB}^{R'S'} X_{R'}^\rho X_S^\sigma Y_\alpha^A, \text{ etc.}$$

Notice also that the objects $\xi_\alpha^{B'}$ and λ_A^β describing the splitting of the twistors can be viewed as the projecting parts of $\xi_\alpha^\beta := \xi^{-1} D_\alpha^\beta \xi$ and $\delta_\alpha^\beta - \xi_\alpha^\beta$, respectively.

D-Curvature

For $f \in \mathcal{E}[w]$ the projecting part of $D_\alpha^\rho f$ is $\frac{1}{p} X_P^\rho D_\alpha^{P'} f = wf$. Although this is 0th order in f , this part of $D_\alpha^\rho f$ behaves like a first order operator because of the weight factor, w . In particular $\frac{1}{p} X_P^\rho D_\alpha^{P'}$ satisfies a Leibniz rule and so therefore so does D_α^ρ . It follows immediately that, acting on $\mathcal{E}^\mu[w]$, $[D_\alpha^\rho, D_\beta^\sigma]$ decomposes into a 0th order curvature part and a 1st order torsion part. In fact it is easy using the identities (4.171) and (4.180) to verify that

$$[D_\alpha^\rho, D_\beta^\sigma] v^\mu = W_{\alpha\beta\gamma}^{\rho\sigma\mu} v^\gamma - W_{\alpha\beta\gamma}^{\rho\sigma\nu} D_\nu^\gamma v^\mu + \delta_\alpha^\sigma D_\beta^\rho v^\mu - \delta_\beta^\rho D_\alpha^\sigma v^\mu,$$

$$\text{where} \quad W_{\alpha\beta\gamma}^{\rho\sigma\mu} = X_{A'}^\rho X_{B'}^\sigma Y_\alpha^A Y_\beta^B W_{AB\gamma}^{A'B'\mu}.$$

4.13.4 Application: Rovelli's Loop Quantum Gravity

4.13.4.1 Introduction to Loop Quantum Gravity

Recall (from subsection 3.10.4 above) that Carlo Rovelli developed (in the last decade of the 20th Century) the so-called loop approach to quantum

gravity (see [Rovelli (1998)] and references therein). The first announcement of this approach was given in [Rovelli and Smolin (1987)]. Together with string theory, this approach provides another serious candidate theory of quantum gravity. It provides a physical picture of Planck scale quantum geometry, calculation techniques, definite quantitative predictions, and a tool for discussing classical problems such as black hole thermodynamics.

String theory and *loop quantum gravity* differ not only because they explore distinct physical hypotheses, but also because they are expressions of two separate communities of scientists, which have sharply distinct prejudices, and view the problem of quantum gravity in surprisingly different manners. As Rovelli says: “I heard the following criticism to loop quantum gravity: ‘Loop quantum gravity is certainly physically wrong, because:

- (1) it is not supersymmetric, and
- (2) is formulated in four dimensions’.

But experimentally, the world still insists on looking four-dimensional and not supersymmetric. In my opinion, people should be careful of not being blinded by their own speculation, and mistaken interesting hypotheses (such as supersymmetry and high-dimensions) for established truth. But string theory may claim extremely remarkable theoretical successes and is today the leading and most widely investigated candidate theory of quantum gravity” [Rovelli (1998)].

High energy physics has obtained spectacular successes during this Century, culminated with the (far from linear) establishment of quantum field theory as the general form of dynamics and with the comprehensive success of the $SU(3) \times SU(2) \times U(1)$ *Standard Model*. Thanks to this success, now a few decades old, physics is in a condition in which it has been very rarely: there are no experimental results that clearly challenge, or clearly escape, the present fundamental theory of the world. The theory we have encompasses virtually everything – except gravitational phenomena. From the point of view of a particle physicist, gravity is then simply the last and weakest of the interactions. It is natural to try to understand its quantum properties using the strategy that has been so successful for the rest of microphysics, or variants of this strategy. The search for a conventional quantum field theory capable of embracing gravity has spanned several decades and, through an adventurous sequence of twists, moments of excitement and disappointments, has lead to string theory. The foundations of string theory are not yet well understood; and it is not yet entirely clear how a supersymmetric theory in 10 or 11 dimensions can be concretely used

for deriving comprehensive univocal predictions about our world.

In string theory, gravity is just one of the excitations of a string (or other extended object) living over some background metric space. The existence of such background metric space, over which the theory is defined, is needed for the formulation and for the interpretation of the theory, not only in perturbative string theory, but in the recent attempts of a non-perturbative definition of the theory, such as M theory, as well, in my understanding. Thus, for a physicist with a high energy background, the problem of quantum gravity is now reduced to an aspect of the problem of understanding what is the mysterious non-perturbative theory that has perturbative string theory as its perturbation expansion, and how to extract information on Planck scale physics from it.

For a *relativist*, on the other hand, the idea of a fundamental description of gravity in terms of physical excitations over a background metric space sounds physically very wrong. The key lesson learned from general relativity is that there is no background metric *over* which physics happens. The world is more complicated than that. Indeed, for a relativist, general relativity is much more than the field theory of a particular force. Rather, it is the discovery that certain classical notions about space and time are inadequate at the fundamental level; they require modifications which are possibly as basics as the ones that quantum mechanics introduced. One of such inadequate notions is precisely the notion of a background metric space (flat or curved), over which physics happens. This profound conceptual shift has led to the understanding of relativistic gravity, to the discovery of black holes, to relativistic astrophysics and to modern cosmology.

From Newton to the beginning of this Century, physics has had a solid foundation in a small number of key notions such as space, time, causality and matter. In spite of substantial evolution, these notions remained rather stable and self-consistent. In the first quarter of this Century, quantum theory and general relativity have modified this foundation in depth. The two theories have obtained solid success and vast experimental corroboration, and can be now considered as established knowledge. Each of the two theories modifies the conceptual foundation of classical physics in a (more or less) internally consistent manner, but we do not have a novel conceptual foundation capable of supporting *both* theories. This is why we do not yet have a theory capable of predicting what happens in the physical regime in which both theories are relevant, the regime of Planck scale phenomena, 10^{-33} cm.

General relativity has taught us not only that space and time share the

property of being dynamical with the rest of the physical entities, but also (more crucially) that space–time location is relational only. Quantum mechanics has taught us that any dynamical entity is subject to Heisenberg’s uncertainty at small scale. Thus, we need *a relational notion of a quantum space–time*, in order to understand Planck scale physics. Thus, for a relativist, the problem of quantum gravity is the problem of bringing a vast conceptual revolution, started with quantum mechanics and with general relativity, to a conclusion and to a new synthesis (see [Rovelli (1997); Smolin (1997)].) In this synthesis, the notions of space and time need to be deeply reshaped, in order to keep into account what we have learned with both our present ‘fundamental’ theories.

Unlike perturbative or non–perturbative string theory, loop quantum gravity is formulated without a background space–time, and is thus a genuine attempt to grasp what is quantum space–time at the fundamental level. Accordingly, the notion of space–time that emerges from the theory is profoundly different from the one on which conventional quantum field theory or string theory are based.

According to Rovelli, the main merit of string theory is that it provides a superbly elegant unification of known fundamental physics, and that it has a well defined perturbation expansion, finite order by order. Its main incompleteness is that its non–perturbative regime is poorly understood, and that we do not have a background–independent formulation of the string theory. In a sense, we do not really know what is the theory we are talking about. Because of this poor understanding of the non perturbative regime of the theory, Planck scale physics and genuine quantum gravitational phenomena are not easily controlled: except for a few computations, there is not much Planck scale physics derived from string theory so far. There are, however, two sets of remarkable physical results. The first is given by some very high energy scattering amplitudes that have been computed. An intriguing aspect of these results is that they indirectly suggest that geometry below the Planck scale cannot be probed –and thus in a sense does not exist– in string theory. The second physical achievement of string theory (which followed the D–branes revolution) is the derivation of the Bekenstein–Hawking black hole entropy formula for certain kinds of black holes.

On the other hand, the main merit of loop quantum gravity is that it provides a well–defined and mathematically rigorous formulation of a background–independent non–perturbative generally covariant quantum field theory. The theory provides a physical picture and quantitative pre-

dictions on the world at the Planck scale. The main incompleteness of the theory regards the dynamics, formulated in several variants. The theory has lead to two main sets of physical results. The first is the derivation of the (Planck scale) eigenvalues of geometrical quantities such as areas and volumes. The second is the derivation of black hole entropy for ‘normal’ black holes (but only up to the precise numerical factor).

The main physical hypotheses on which loop quantum gravity relies are only general relativity and quantum mechanics. In other words, loop quantum gravity is a rather conservative ‘quantization’ of general relativity, with its traditional matter couplings. In this sense, it is very different from string theory, which is based on a strong physical hypothesis with no direct experimental support ‘*that the world is made by strings*’.

Finally, strings and loop gravity, may not necessarily be competing theories: there might be a sort of complementarity, at least methodological, between the two. This is due to the fact that the open problems of string theory regard its background-independent formulation, and loop quantum gravity is precisely a set of techniques for dealing non-perturbatively with background independent theories. Perhaps the two approaches might even, to some extent, converge. Undoubtedly, there are similarities between the two theories: first of all the obvious fact that both theories start with the idea that the relevant excitations at the Planck scale are one dimensional objects – call them loops or strings. [Smolin (1997)] also explored the possible relations between string theory and loop quantum gravity.

Loop quantum gravity is a quantum field theory on a differentiable 4-manifold. We have learned with general relativity that the space-time metric and the gravitational field are the same physical entity. Thus, a quantum theory of the gravitational field is a quantum theory of the space-time metric as well. It follows that quantum gravity cannot be formulated as a quantum field theory over a *metric manifold*, because there is no (classical) metric manifold whatsoever in a regime in which gravity (and therefore the metric) is a quantum variable [Rovelli (1998)].

One could conventionally split the space-time metric into two terms: one to be consider as a background, which gives a metric structure to space-time; the other to be treated as a fluctuating quantum field. This, indeed, is the procedure on which old perturbative quantum gravity, perturbative strings, as well as current non-perturbative string theories (M-theory), are based. In following this path, one assumes, for instance, that the causal structure of space-time is determined by the underlying background metric alone, and not by the full metric. Contrary to this, in loop quantum

gravity we assume that the identification between the gravitational field and the metric-causal structure of space-time holds, and must be taken into account, in the quantum regime as well. Thus, no split of the metric is made, and there is no background metric on space-time.

We can still describe space-time as a (differentiable) manifold (a space without metric structure), over which quantum fields are defined. A classical metric structure will then be defined by expectation values of the gravitational field operator. Thus, the problem of quantum gravity is the problem of understanding what is *a quantum field theory on a manifold*, as opposed to quantum field theory on a metric space. This is what gives quantum gravity its distinctive flavor, so different than ordinary quantum field theory. In all versions of ordinary quantum field theory, the metric of space-time plays an essential role in the construction of the basic theoretical tools (creation and annihilation operators, canonical commutation relations, gaussian measures, propagators); these tools cannot be used in quantum field over a manifold.

Technically, the difficulty due to the absence of a background metric is circumvented in loop quantum gravity by defining the quantum theory as a representation of a Poisson algebra of classical observables, which can be defined without using a background metric. The idea that the quantum algebra at the basis of quantum gravity is not the canonical commutation relation algebra, but the Poisson algebra of a different set of observables has long been advocated by [Isham (1984)], whose ideas have been very influential in the birth of loop quantum gravity. The algebra on which loop gravity is the loop algebra [Rovelli and Smolin (1990)].

In choosing the *loop algebra* as the basis for the quantization, we are essentially assuming that Wilson loop operators are well defined in the Hilbert space of the theory. In other words, that certain states concentrated on one dimensional structures (loops and graphs) have finite norm. This is a subtle non trivial assumptions entering the theory. It is the key assumption that characterizes loop gravity. If the approach turned out to be wrong, it will likely be because this assumption is wrong. The Hilbert space resulting from adopting this assumption is not a Fock space. Physically, the assumption corresponds to the idea that quantum states can be decomposed on a basis of *Faraday lines*-excitations (as Minkowski QFT states can be decomposed on a particle basis).

Furthermore, this is an assumption that *fails* in conventional quantum field theory, because in that context well defined operators and finite norm states need to be smeared in at least three dimensions, and 1D objects are

too singular. The fact that at the basis of loop gravity there is a mathematical assumption that fails for *conventional Yang–Mills quantum field theory* is probably at the origin of some of the resistance that loop quantum gravity encounters among some high energy theorists. What distinguishes gravity from Yang–Mills (YM) theories, however, and makes this assumption viable in gravity even if it fails for YM theory is *diffeomorphism invariance*. The loop states are singular states that span a ‘huge’ non-separable state space. Non-perturbative diffeomorphism invariance plays two roles. First, it wipes away the infinite redundancy. Second, it ‘smears’ a loop state into a knot state, so that the physical states are not really concentrated in one dimension, but are, in a sense, smeared all over the entire manifold by the non-perturbative diffeomorphisms [Rovelli (1998)].

Conventional field theories are not invariant under a diffeomorphism acting on the dynamical fields. Every field theory, suitably formulated, is trivially invariant under a diffeomorphism acting on *everything*. General relativity, on the contrary is invariant under such transformations. More precisely, every general relativistic theory has this property. Thus, diffeomorphism invariance is not a feature of just the gravitational field: it is a feature of physics, once the existence of relativistic gravity is taken into account. Thus, one can say that the gravitational field is not particularly ‘special’ in this regard, but that diffeomorphism invariance is a property of the physical world that can be disregarded only in the approximation in which the dynamics of gravity is neglected.

Now, *diffeomorphism invariance* is the technical implementation of a physical idea, due to Einstein. The idea is a deep modification of the pre-general-relativistic (pre-GR) notions of space and time. In pre-GR physics, we assume that physical objects can be localized in space and time with respect to a fixed non-dynamical background structure. Operationally, this background space–time can be defined by means of physical reference-system objects, but these objects are considered as dynamically decoupled from the physical system that one studies. This conceptual structure fails in a relativistic gravitational regime. In general relativistic physics, the physical objects are localized in space and time only with respect to each other. Therefore if we ‘displace’ all dynamical objects in space–time at once, we are not generating a different state, but an equivalent mathematical description of the same physical state. Hence, diffeomorphism invariance.

Accordingly, a physical state in GR is not ‘located’ somewhere. Pictorially, GR is not physics over a stage, it is the dynamical theory of (or including) the stage itself. Loop quantum gravity is an attempt to imple-

ment this subtle relational notion of space–time localization in quantum field theory. In particular, the basic quantum field theoretical excitations cannot be localized somewhere as, say, photons are. They are quantum excitations of the ‘stage’ itself, not excitations over a stage. Intuitively, one can understand from this discussion how knot theory plays a role in the theory. First, we define quantum states that correspond to loop–like excitations of the gravitational field, but then, when factoring away diffeomorphism invariance, the location of the loop becomes irrelevant. The only remaining information contained in the loop is then its knotting (a knot is a loop up to its location). Thus, diffeomorphism invariant physical states are labelled by knots. A knot represent an elementary quantum excitation of space. It is not here or there, since it *is* the space with respect to which here and there can be defined. A knot state is an elementary quantum of space. In this manner, loop quantum gravity ties the new notion of space and time introduced by general relativity with quantum mechanics.

4.13.4.2 Formalism of Loop Quantum Gravity

The starting point is classical general relativity formulated in terms of the *Ashtekar phase–space formalism* (see [Ashtekar (1991)]). Recall that classical general relativity can be formulated in the phase–space form as follows.

We fix a 3D manifold M (compact and without boundaries) and consider a smooth real $SU(2)$ –connection $A_a^i(x)$ and a vector density $\tilde{E}_i^a(x)$ (transforming in the vector representation of $SU(2)$) on M . We use $a, b, \dots = 1, 2, 3$ for spatial indices and $i, j, \dots = 1, 2, 3$ for internal indices. The internal indices can be viewed as labelling a basis in the Lie algebra of $SU(2)$ or the three axis of a local triad. We indicate coordinates on M with x . The relation between these fields and conventional metric gravitational variables is as follows: $\tilde{E}_i^a(x)$ is the (densitized) inverse triad, related to the 3D metric $g_{ab}(x)$ of constant–time surfaces by

$$g^{ab} = \tilde{E}_i^a \tilde{E}_i^b, \quad (4.181)$$

where g is the determinant of g_{ab} ; and

$$A_a^i(x) = \Gamma_a^i(x) + \gamma k_a^i(x); \quad (4.182)$$

$\Gamma_a^i(x)$ is the *spin connection* associated to the triad, (defined by $\partial_{[a} e_{b]}^i = \Gamma_{[a}^i e_{b]}^i$, where e_a^i is the triad).

$k_a^i(x)$ is the extrinsic curvature of the constant time three surface.

In (4.182), γ is a constant, denoted the Immirzi parameter, that can be chosen arbitrarily (it will enter the Hamiltonian constraint). Different choices for γ yield different versions of the formalism, all equivalent in the classical domain. If we choose γ to be equal to the imaginary unit, $\gamma = \sqrt{-1}$, then A is the standard *Ashtekar connection*, which can be shown to be the projection of the self-dual part of the 4D spin connection on the constant time surface. If we choose $\gamma = 1$, we get the real Barbero connection. The Hamiltonian constraint of Lorentzian general relativity has a particularly simple form in the $\gamma = \sqrt{-1}$ formalism, while the Hamiltonian constraint of Euclidean general relativity has a simple form when expressed in terms of the $\gamma = 1$ real connection. Other choices of γ are viable as well. In particular, it has been argued that the quantum theory based on different choices of γ are genuinely physical inequivalent, because they yield ‘geometrical quanta’ of different magnitude [Rovelli (1998)]. Apparently, there is a unique choice of γ yielding the correct $1/4$ coefficient in the *Bekenstein–Hawking formula*.

The spinorial version of the Ashtekar variables is given in terms of the Pauli matrices $\sigma_i, i = 1, 2, 3$, or the $\mathfrak{su}(2)$ generators $\tau_i = -\frac{i}{2} \sigma_i$, by

$$\tilde{E}^a(x) = -i \tilde{E}_i^a(x) \sigma_i = 2\tilde{E}_i^a(x) \tau_i, \quad (4.183)$$

$$A_a(x) = -\frac{i}{2} A_a^i(x) \sigma_i = A_a^i(x) \tau_i. \quad (4.184)$$

Thus, $A_a(x)$ and $\tilde{E}^a(x)$ are 2×2 anti-Hermitian complex matrices.

The theory is invariant under local $SU(2)$ gauge transformations, three-dimensional diffeomorphisms of the manifold on which the fields are defined, as well as under (coordinate) time translations generated by the Hamiltonian constraint. The full dynamical content of general relativity is captured by the three constraints that generate these gauge invariances (see [Ashtekar (1991)]).

4.13.4.3 Loop Algebra

Certain classical quantities play a very important role in the quantum theory. These are: the trace of the holonomy of the connection, which is labelled by loops on the three manifold; and the higher order loop variables, obtained by inserting the E field (in n distinct points, or ‘hands’) into the holonomy trace. More precisely, given a loop α in M and the points

$s_1, s_2, \dots, s_n \in \alpha$ we define:

$$\mathcal{T}[\alpha] = -\text{Tr}[U_\alpha], \quad (4.185)$$

$$\mathcal{T}^a[\alpha](s) = -\text{Tr}[U_\alpha(s, s)\tilde{E}^a(s)] \quad (4.186)$$

and, in general

$$\mathcal{T}^{a_1 a_2}[\alpha](s_1, s_2) = -\text{Tr}[U_\alpha(s_1, s_2)\tilde{E}^{a_2}(s_2)U_\alpha(s_2, s_1)\tilde{E}^{a_1}(s_1)], \quad (4.187)$$

$$\mathcal{T}^{a_1 \dots a_N}[\alpha](s_1 \dots s_N) = -\text{Tr}[U_\alpha(s_1, s_N)\tilde{E}^{a_N}(s_N)U_\alpha(s_N, s_{N-1}) \dots \tilde{E}^{a_1}(s_1)]$$

where $U_\alpha(s_1, s_2) \sim \mathcal{P} \exp\{\int_{s_1}^{s_2} A_a(\alpha(s))ds\}$ is the parallel propagator of A_a along α , defined by

$$\frac{d}{ds}U_\alpha(1, s) = \frac{d\alpha_a(s)}{ds} A_a(\alpha(s)) U_\alpha(1, s). \quad (4.188)$$

These are the loop observables, previously introduced in YM theories.

The loop observables coordinate the phase space and have a closed Poisson algebra, denoted the loop algebra. This algebra has a remarkable geometrical flavor. For instance, the Poisson bracket between $\mathcal{T}[\alpha]$ and $\mathcal{T}^a[\beta](s)$ is non vanishing only if $\beta(s)$ lies over α ; if it does, the result is proportional to the holonomy of the Wilson loops obtained by joining α and β at their intersection (by rerouting the 4 legs at the intersection). More precisely

$$\{\mathcal{T}[\alpha], \mathcal{T}^a[\beta](s)\} = \Delta^a[\alpha, \beta(s)] [\mathcal{T}[\alpha\#\beta] - \mathcal{T}[\alpha\#\beta^{-1}]]. \quad (4.189)$$

Here

$$\Delta^a[\alpha, x] = \int ds \frac{d\alpha^a(s)}{ds} \delta^3(\alpha(s), x) \quad (4.190)$$

is a vector distribution with support on α and $\alpha\#\beta$ is the loop obtained starting at the intersection between α and β , and following first α and then β . β^{-1} is β with reversed orientation.

A (non- $SU(2)$ gauge invariant) quantity that plays a role in certain aspects of the theory, particularly in the regularization of certain operators, is obtained by integrating the E field over a two dimensional surface S

$$E[S, f] = \int_S dS_a \tilde{E}_i^a f^i, \quad (4.191)$$

where f is a function on the surface S , taking values in the Lie algebra of $SU(2)$. In alternative to the full loop observables (4.185, 4.186, 4.187), one also can take the holonomies and $E[S, f]$ as elementary variables.

4.13.4.4 Loop Quantum Gravity

The kinematic of a quantum theory is defined by an algebra of ‘elementary’ operators (such as x and $i\hbar d/dx$, or creation and annihilation operators) on a Hilbert space \mathcal{H} . The physical interpretation of the theory is based on the connection between these operators and classical variables, and on the interpretation of \mathcal{H} as the space of the quantum states. The dynamics is governed by a Hamiltonian, or, as in general relativity, by a set of quantum constraints, constructed in terms of the elementary operators. To assure that the quantum Heisenberg equations have the correct classical limit, the algebra of the elementary operator has to be isomorphic to the Poisson algebra of the elementary observables. This yields the *heuristic quantization rule*: ‘promote Poisson brackets to commutators’. In other words, define the quantum theory as a linear representation of the Poisson algebra formed by the elementary observables. The kinematics of the quantum theory is defined by a unitary representation of the loop algebra.

We can start *à la Schrödinger*, by expressing quantum states by means of the amplitude of the connection, namely by means of functionals $\Psi(A)$ of the (smooth) connection. These functionals form a linear space, which we promote to a Hilbert space by defining a inner product. To define the inner product, we choose a particular set of states, which we denote ‘cylindrical states’ and begin by defining the scalar product between these.

Pick a graph Γ , say with n links, denoted $\gamma_1 \dots \gamma_n$, immersed in the manifold M . For technical reasons, we require the links to be analytic. Let $U_i(A) = U_{\gamma_i}$, $i = 1, \dots, n$ be the parallel transport operator of the connection A along γ_i . $U_i(A)$ is an element of $SU(2)$. Pick a function $f(g_1 \dots g_n)$ on $[SU(2)]^n$. The graph Γ and the function f determine a functional of the connection as follows

$$\psi_{\Gamma, f}(A) = f(U_1(A), \dots, U_n(A)), \quad (4.192)$$

(these states are called cylindrical states because they were previously introduced as cylindrical functions for the definition of a cylindrical measure). Notice that we can always ‘enlarge the graph’, in the sense that if Γ is a subgraph of Γ' we can write

$$\psi_{\Gamma, f}(A) = \psi_{\Gamma', f'}(A), \quad (4.193)$$

by simply choosing f' independent from the U_i ’s of the links which are in Γ' but not in Γ . Thus, given any two cylindrical functions, we can always view them as having the same graph (formed by the union of the two

graphs). Given this observation, we define the scalar product between any two cylindrical functions, by

$$(\psi_{\Gamma,f}, \psi_{\Gamma,h}) = \int_{SU(2)^n} dg_1 \dots dg_n \overline{f(g_1 \dots g_n)} h(g_1 \dots g_n). \quad (4.194)$$

where dg is the Haar measure on $SU(2)$. This scalar product extends by linearity to finite linear combinations of cylindrical functions. It is not difficult to show that (4.194) defines a well defined scalar product on the space of these linear combinations. Completing the space of these linear combinations in the Hilbert norm, we get a Hilbert space \mathcal{H} . This is the (unconstrained) quantum state space of loop gravity.²⁴ \mathcal{H} carries a natural unitary representation of the diffeomorphism group and of the group of the local $SU(2)$ transformations, obtained transforming the argument of the functionals. An important property of the scalar product (4.194) is that it is invariant under both these transformations.

\mathcal{H} is non-separable. At first sight, this may seem as a serious obstacle for its physical interpretation. But we will see below that after factoring away diffeomorphism invariance we may get a separable Hilbert space. Also, standard spectral theory holds on \mathcal{H} , and it turns out that using spin networks (discussed below) one can express \mathcal{H} as a direct sum over finite dimensional subspaces which have the structure of Hilbert spaces of spin systems; this makes practical calculations very manageable.

Finally, we will use a Dirac notation and write

$$\Psi(A) = \langle A | \Psi \rangle, \quad (4.195)$$

in the same manner in which one may write $\psi(x) = \langle x | \Psi \rangle$ in ordinary quantum mechanics. As in that case, however, we should remember that $|A\rangle$ is not a normalizable state.

4.13.4.5 Loop States and Spin Network States

A subspace \mathcal{H}_0 of \mathcal{H} is formed by states invariant under $SU(2)$ gauge transformations. We now define an orthonormal basis in \mathcal{H}_0 . This ba-

²⁴This construction of \mathcal{H} as the closure of the space of the cylindrical functions of smooth connections in the scalar product (4.194) shows that \mathcal{H} can be defined without the need of recurring to C^* algebraic techniques, distributional connections or the Ashtekar-Lewandowski measure. The casual reader, however, should be warned that the resulting \mathcal{H} topology is different than the natural topology on the space of connections: if a sequence Γ_n of graphs converges point-wise to a graph Γ , the corresponding cylindrical functions $\psi_{\Gamma_n,f}$ do not converge to $\psi_{\Gamma,f}$ in the \mathcal{H} Hilbert space topology.

sis represents a very important tool for using the theory. It was introduced in [Rovelli and Smolin (1995)] and developed in [Baez (1996a); Baez (1996b)]; it is denoted *spin network* basis.

First, given a loop α in M , there is a normalized state $\psi_\alpha(A)$ in \mathcal{H} , which is obtained by taking $\Gamma = \alpha$ and $f(g) = -\text{Tr}(g)$. Namely

$$\psi_\alpha(A) = -\text{Tr}(U_\alpha(A)). \quad (4.196)$$

We introduce a Dirac notation for the abstract states, and denote this state as $|\alpha\rangle$. These states are called loop states. Using Dirac notation, we can write

$$\psi_\alpha(A) = \langle A | \alpha \rangle, \quad (4.197)$$

It is easy to show that loop states are normalizable. Products of loop states are normalizable as well. Following tradition, we denote with α also a multi-loop, namely a collection of (possibly overlapping) loops $\{\alpha_1, \dots, \alpha_n\}$, and we call

$$\psi_\alpha(A) = \psi_{\alpha_1}(A) \times \dots \times \psi_{\alpha_n}(A) \quad (4.198)$$

– a multi-loop state. Multi-loop states represented the main tool for loop quantum gravity before the discovery of the spin network basis. Linear combinations of multi-loop states over-span \mathcal{H} , and therefore a generic state $\psi(A)$ is fully characterized by its projections on the multi-loop states, namely by

$$\psi(\alpha) = (\psi_\alpha, \psi). \quad (4.199)$$

The ‘old’ loop representation was based on representing quantum states in this manner, namely by means of the functionals $\psi(\alpha)$ over loop space defined in (4.199).

Next, consider a graph Γ . A ‘coloring’ of Γ is given by the following.

- (1) Associate an irreducible representation of $SU(2)$ to each link of Γ . Equivalently, we may associate to each link γ_i a half integer number s_i , the spin of the irreducible, or, equivalently, an integer number p_i , the ‘color’ $p_i = 2s_i$.
- (2) Associate an invariant tensor v in the tensor product of the representations $s_1 \dots s_n$, to each node of Γ in which links with spins $s_1 \dots s_n$ meet. An invariant tensor is an object with n indices in the representations $s_1 \dots s_n$ that transform covariantly. If $n = 3$, there is only one invariant tensor (up to a multiplicative factor), given by the Clebsh–Gordon

coefficient. An invariant tensor is also called an *intertwining tensor*. All invariant tensors are given by the standard Clebsch–Gordon theory. More precisely, for fixed $s_1 \dots s_n$, the invariant tensors form a finite dimensional linear space. Pick a basis v_j in this space, and associate one of these basis elements to the node. Notice that invariant tensors exist only if the tensor product of the representations $s_1 \dots s_n$ contains the trivial representation. This yields a condition on the coloring of the links. For $n = 3$, this is given by the well known Clebsch–Gordan condition: each color is not larger than the sum of the other two, and the sum of the three colors is even.

We indicate a colored graph by $\{\Gamma, \vec{s}, \vec{v}\}$, or simply $S = \{\Gamma, \vec{s}, \vec{v}\}$, and denote it a ‘spin network’. (It was R. Penrose who first had the intuition that this mathematics could be relevant for describing the quantum properties of the geometry, and who gave the first version of *spin network theory* [Penrose (1971a); Penrose (1971b)].)

Given a spin network S , we can construct a state $\Psi_S(A)$ as follows. We take the propagator of the connection along each link of the graph, in the representation associated to that link, and then, at each node, we contract the matrices of the representation with the invariant tensor. We get a state $\Psi_S(A)$, which we also write as

$$\psi_S(A) = \langle A | S \rangle. \quad (4.200)$$

One can then show the following.

- The spin network states are normalizable. The normalization factor is computed in [DePietri and Rovelli (1996)].
- They are $SU(2)$ gauge invariant.
- Each spin network state can be decomposed into a finite linear combination of products of loop states.
- The (normalized) spin network states form an orthonormal basis for the gauge $SU(2)$ invariant states in \mathcal{H} (choosing the basis of invariant tensors appropriately).
- The scalar product between two spin network states can be easily computed graphically and algebraically.

The spin network states provide a very convenient basis for the quantum theory.

The spin network states defined above are $SU(2)$ gauge invariant. There exists also an extension of the spin network basis to the full Hilbert space.

4.13.4.6 Diagrammatic Representation of the States

A diagrammatic representation for the states in \mathcal{H} is very useful in concrete calculations. First, associate to a loop state $|\alpha\rangle$ a diagram in M , formed by the loop α itself. Next, notice that we can *multiply* two loop states, obtaining a normalizable state. We represent the product of n loop states by the diagram formed by the set of the n (possibly overlapping) corresponding loops (we denote this set ‘multi-loop’). Thus, linear combinations of multi-loops diagrams represent states in \mathcal{H} . Representing states as linear combinations of multi-loops diagrams makes computation in \mathcal{H} easy.

Now, the spin network state defined by the graph with no nodes α , with color 1, is clearly, by definition, the loop state $|\alpha\rangle$, and we represent it by the diagram α . The spin network state $|\alpha, n\rangle$ determined by the graph without nodes α , with color n can be obtained as follows. Draw n parallel lines along the loop α ; cut all lines at an arbitrary point of α , and consider the $n!$ diagrams obtained by joining the legs after a permutation. The linear combination of these $n!$ diagrams, taken with alternate signs (namely with the sign determined by the parity of the permutation) is precisely the state $|\alpha, n\rangle$. The reason of this key result can be found in the fact that an irreducible representation of $SU(2)$ can be obtained as the totally symmetric tensor product of the fundamental representation with itself (for details, see [DePietri and Rovelli (1996)]).

Next, consider a graph Γ with nodes. Draw n_i parallel lines along each link γ_i . Join pairwise the end points of these lines at each node (in an arbitrary manner), in such a way that each line is joined with a line from a different link. In this manner, one get a multi-loop diagram. Now antisymmetrize the n_i parallel lines along each link, obtaining a linear combination of diagrams representing a state in \mathcal{H} . One can show that this state is a spin network state, where n_i is the color of the links and the color of the nodes is determined by the pairwise joining of the legs chosen [DePietri and Rovelli (1996)]. Again, simple $SU(2)$ representation theory is behind this result.

More in detail, if a node is trivalent (has 3 adjacent links), then we can join legs pairwise only if the total number of the legs is even, and if the number of the legs in each link is smaller or equal than the sum of the number of the other two. This can be immediately recognized as the *Clebsch–Gordan condition*. If these conditions are satisfied, there is only a single way of joining legs. This corresponds to the fact that there is only one invariant tensor in the product of three irreducible of $SU(2)$. Higher

valence nodes can be decomposed in trivalent ‘virtual’ nodes, joined by ‘virtual’ links. Orthogonal independent invariant tensors are obtained by varying over all allowed colorings of these virtual links (compatible with the Clebsch–Gordan conditions at the virtual nodes). Different decompositions of the node give different orthogonal bases. Thus the total (links and nodes) coloring of a spin network can be represented by means of the coloring of the real and the virtual links (see Figure 4.10).

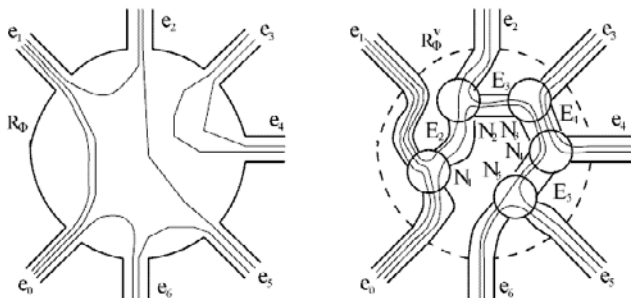


Fig. 4.10 Construction of ‘virtual’ nodes links over an n -valent node in a graph Γ .

Viceversa, multi-loop states can be decomposed in spin network states by simply symmetrizing along (real and virtual) nodes. This can be done particularly easily diagrammatically, as illustrated by the graphical formulae in [Rovelli and Smolin (1995); DePietri and Rovelli (1996)]. These are standard formulae. In fact, it is well known that the tensor algebra of the $SU(2)$ irreducible representations admits a completely graphical notation. This graphical notation has been widely used for instance in nuclear and atomic physics. The application of this diagrammatic calculus to quantum gravity is described in detail in [DePietri and Rovelli (1996)].

4.13.4.7 Quantum Operators

Now, we define the *quantum operators*, corresponding to the \mathcal{T} -variables, as linear operators on \mathcal{H} . These form a representation of the loop variables Poisson algebra. The operator $\mathcal{T}[\alpha]$ acts diagonally

$$\mathcal{T}[\alpha]\Psi(A) = -\text{Tr } U_\alpha(A) \Psi(A),$$

(recall that products of loop states and spin-network states are normalizable states). In diagrammatic notation, the operator simply adds a loop to

a (linear combination of) multi-loops

$$\mathcal{T}[\alpha] |\Psi\rangle = |\alpha\rangle |\Psi\rangle.$$

Higher order loop operators are expressed in terms of the elementary ‘grasp’ operation. Consider first the operator $\mathcal{T}^a(s)[\alpha]$, with one hand in the point $\alpha(s)$. The operator annihilates all loop states that do not cross the point $\alpha(s)$. Acting on a loop state $|\beta\rangle$, it gives

$$\mathcal{T}^a(s)[\alpha] |\beta\rangle = l_0^2 \Delta^a[\beta, \alpha(s)] [|\alpha\#\beta\rangle - |\alpha\#\beta^{-1}\rangle], \quad (4.201)$$

where we have introduced the elementary length l_0 by

$$l_0^2 = \hbar G = \frac{16\pi\hbar G_{\text{Newton}}}{c^3} = 16\pi l_{\text{Planck}}^2 \quad (4.202)$$

and Δ^a and $\#$ were defined above. This action extends by linearity, continuity and by the Leibniz rule to products and linear combinations of loop states, and to the full \mathcal{H} . In particular, it is not difficult to compute its action on a spin network state [DePietri and Rovelli (1996)]. Higher order loop operators act similarly. It is easy to verify that these operators provide a representation of the classical Poisson loop algebra.

All the operators in the theory are then constructed in terms of these basics loop operators, in the same way in which in conventional QFT one constructs all operators, including the Hamiltonian, in terms of creation and annihilation operators. The construction of the composite operators requires the development of regularization techniques that can be used in the absence of a background metric.

4.13.4.8 Loop v.s. Connection Representation

Imagine we want to quantize the one dimensional harmonic oscillator. We can consider the Hilbert space of square integrable functions $\psi(x)$ on the real line, and express the momentum and the Hamiltonian as differential operators. Denote the eigenstates of the Hamiltonian as $\psi_n(x) = \langle x|n\rangle$. It is well known that the theory can be expressed entirely in algebraic form in terms of the states $|n\rangle$. In doing so, all elementary operators are algebraic: $\hat{x}|n\rangle = \frac{1}{\sqrt{2}}(|n-1\rangle + (n+1)|n+1\rangle)$, $\hat{p}|n\rangle = \frac{-i}{\sqrt{2}}(|n-1\rangle - (n+1)|n+1\rangle)$. Similarly, in quantum gravity we can directly construct the quantum theory in the spin-network (or loop) basis, without ever mentioning functionals of the connections. This representation of the theory is denoted the *loop representation*.

A section of the first paper on loop quantum gravity by [Rovelli and Smolin (1990)] was devoted to a detailed study of ‘transformation theory’ (in the sense of Dirac) on the state space of quantum gravity, and in particular on the relations between the loop states

$$\psi(\alpha) = \langle \alpha | \psi \rangle \quad (4.203)$$

and the states $\psi(A)$ giving the amplitude for a connection field configuration A , and defined by

$$\psi(A) = \langle A | \psi \rangle. \quad (4.204)$$

Here $|A\rangle$ are ‘eigenstates of the connection operator’, or, more precisely (since the operator corresponding to the connection is ill defined in the theory) the generalized states that satisfy

$$T[\alpha] |A\rangle = -\text{Tr}[\mathcal{P}e^{\int_\alpha A}] |A\rangle. \quad (4.205)$$

However, at the time of [Rovelli and Smolin (1990)] the lack of a scalar product made transformation theory quite involved.

On the other hand, the introduction of the scalar product (4.194) gives a rigorous meaning to the loop transform. In fact, we can write, for every spin network S , and every state $\psi(A)$

$$\psi(S) = \langle S | \psi \rangle = (\psi_S, \psi). \quad (4.206)$$

This equation defines a unitary mapping between the two presentations of \mathcal{H} : the ‘loop representation’, in which one works in terms of the basis $|S\rangle$; and the ‘connection representation’, in which one uses wave functionals $\psi(A)$.

The development of the connection representation followed a winding path through C^* -algebraic and measure theoretical methods. The work of [DePietri (1997)] has proven the unitary equivalence of the two formalisms.

4.14 Application: Seiberg–Witten Monopole Field Theory

Some of the most important physical problems of contemporary theoretical physics concern the behavior of gauge theories and string theory at strong coupling. For gauge theories, these include the problems of confinement of color, of dynamical chiral symmetry breaking, of the strong coupling behavior of chiral gauge theories, of the dynamical breaking of supersymmetry. In each of these areas, major advances have been achieved over the

past few years, and a useful resolution of some of these difficult problems appears to be within sight. For string theory, these include the problems of dynamical compactification of the 10D theory to string vacua with 4D and of supersymmetry breaking at low energies. Already, it has become clear that, at strong coupling, the string spectrum is radically altered and effectively derives from the unique 11D M-theory.

This rapid progress was driven in large part by the Seiberg–Witten (SW) solution of $N = 2$ supersymmetric Yang–Mills (YM) theory for $SU(2)$ gauge group [Seiberg and Witten (1994a); Seiberg and Witten (1994b)] and by the discovery of D -branes in string theory. Some of the key ingredients underlying these developments are [D'Hoker and Phong (1998a); D'Hoker and Phong (1997)]:

- Restriction to solving for the low energy behavior of the non-perturbative dynamics, summarized by the low energy effective action of the theory.
- High degrees of supersymmetry. This has the effect of imposing certain *holomorphicity constraints* on parts of the low energy effective action, and thus of restricting its form considerably. For gauge theories in 4D, the following degrees of supersymmetry can be distinguished:
 - (1) $N = 1$ supersymmetry supports chiral fermions and is the starting point for the Minimal Supersymmetric Standard Model, the simplest extension of the Standard Model to include supersymmetric partners.
 - (2) $N = 2$ supersymmetry only supports non-chiral fermions and is thus less realistic as a particle physics model, but appears better 'solvable'. This is where the SW-solution was constructed.
 - (3) $N = 4$ is the maximal amount of supersymmetry, and a special case of $N = 2$ supersymmetry with only non-chiral fermions and vanishing renormalization group β -function. Dynamically, the latter theory is the simplest amongst 4D gauge theories, and offers the best hopes for admitting an exact solution.
- Electric-magnetic and *Montonen–Olive duality* [Montonen and Olive (1977)]. Recall that the free Maxwell equations are invariant under electric-magnetic duality when $\vec{E} \rightarrow \vec{B}$ and $\vec{B} \rightarrow -\vec{E}$. In the presence of matter, duality will require the presence of both electric charge e and magnetic monopole charge g whose magnitude is related by Dirac quantization: $e \cdot g \sim \hbar$. Thus, weak electric coupling is related to large magnetic coupling. Conversely, problems of large electric coupling

(such as confinement of the color electric charge of quarks) are mapped by duality into problems of weak magnetic charge. It was conjectured by [Montonen and Olive (1977)] that the $N = 4$ supersymmetric YM theory for any gauge algebra \mathfrak{g} is mapped under the interchange of electric and magnetic charges, i.e., under $e \leftrightarrow 1/e$ into the theory with dual gauge algebra \mathfrak{g}^\vee . When combined with the shift-invariance of the instanton angle θ this symmetry is augmented to the duality group $SL(2, \mathbb{Z})$, or a subgroup thereof.

Of central interest to many of these exciting developments is the 4D supersymmetric YM theory with maximal supersymmetry, $N = 4$, and with arbitrary gauge algebra \mathfrak{g} . As an $N = 2$ supersymmetric theory, the theory has a \mathfrak{g} -gauge multiplet, and a hypermultiplet in the adjoint representation of \mathfrak{g} with mass m .

This generalized theory enjoys many of the same properties as the $N = 4$ theory: it has the same field contents; it is ultra-violet finite; it has vanishing renormalization group β -function, and it is expected to have Montonen–Olive duality symmetry. For vanishing hypermultiplet mass $m = 0$, the $N = 4$ theory is recovered. For $m \rightarrow \infty$, it is possible to choose dependences of the gauge coupling and of the gauge scalar expectation values so that the limiting theory is one of many interesting $N = 2$ supersymmetric YM theories. Amongst these possibilities for $\mathfrak{g} = SU(N)$ for example, are the theories with any number of hypermultiplets in the fundamental representation of $SU(N)$, or with product gauge algebras $SU(N_1) \times SU(N_2) \times \cdots \times SU(N_p)$, and hypermultiplets in fundamental and bi-fundamental representations of these product algebras.

An outstanding problem in non-Abelian gauge theory has been to make reliable predictions about the (non-perturbative) strong coupling region. N. Seiberg and E. Witten in [Seiberg and Witten (1994a); Seiberg and Witten (1994b)] studied $N = 2$ supersymmetric gauge theories in 4D with matter multiplets. For all such models for which the gauge group is $SU(2)$, they derived the exact metric on the *moduli space* of quantum vacua and the exact spectrum of the stable massive states. Seiberg and Witten have shown that the local part of the effective action is governed by a single analytic function \mathcal{F} of a complex variable; they made an Ansatz for the \mathcal{F} that satisfies all the physical criteria and embodies electromagnetic duality, thus directly connecting the weak to the strong coupling regions. A number of new physical phenomena occurred, such as chiral symmetry breaking that was driven by the condensation of magnetic monopoles that carried

global quantum numbers. For those cases in which conformal invariance was broken only by mass terms, their formalism automatically gave results that were invariant under electric–magnetic duality.

4.14.1 *SUSY Formalism*

4.14.1.1 *$N = 2$ Supersymmetry*

First, recall the essentials of the $N = 2$ supersymmetry algebra (for details see, e.g., [West (1990)]). The algebra is given by [Flume *et. al.* (1996)]

$$\{Q_\alpha^i, \bar{Q}_\beta^k\} = \delta_{ik} \sigma_{\alpha\beta}^\mu P_\mu, \quad \{Q_\alpha^i, Q_\beta^k\} = \epsilon_{ik} \epsilon_{\alpha\beta} Z,$$

plus the Hermitian conjugate of the second relation, where $i, k = 1, 2$ and Z is a central charge. This algebra is realized on the simplest possible non-trivial super-multiplet: $\Psi \supset \{\phi, \psi, A_\mu; \mathcal{F}, \mathcal{D}\}$, where ϕ is a complex scalar field, ψ is a Dirac spinor, A_μ is a gauge-field, while \mathcal{F} and \mathcal{D} are complex and real dummy-fields respectively. This $N = 2$ superfield actually consists of two $N = 1$ superfields: $\Phi \supset \{\phi, q, \mathcal{F}\}$ and $V \supset \{A_\mu, f, \mathcal{D}\}$ or $W_\alpha \supset \{\mathcal{F}_{\mu\nu}, f, \mathcal{D}\}$, where ϕ and V/W_α are chiral and vector multiplets respectively, the q and f fields being *Weyl spinors* of opposite chirality. Since the gauge-field A_μ belongs to the adjoint representation of the gauge group G and all the fields belong to the same multiplet, they must all belong to the adjoint representation of G .

4.14.1.2 *$N = 2$ Super-Action*

The super-action for the $N = 2$ superfield just described is given by [Flume *et. al.* (1996)]

$$\mathcal{A} = \text{Im Tr} \int d^4x d^2\theta_\alpha d^2\bar{\theta}_\beta (\Psi)^2.$$

On expanding this action in terms of the $N = 1$ superfields, it becomes

$$\mathcal{A} = \text{Im} \int d^4x d^2\theta_\alpha d^2\bar{\theta}_\beta \left(\bar{A} e^{-2g_o V} A \right) + \tau_o \int d^4x d^2\theta_\alpha (W_\alpha W_\alpha), \quad (4.207)$$

$$\text{where} \quad \tau_o = \frac{\theta_o}{2\pi} + \frac{4\pi i}{g_o^2},$$

the parameter g_o is the usual gauge-coupling constant and θ_o is the *QCD*-vacuum-angle. The exponential in the first term is just the supersymmetric generalization of the covariant derivative. Expanding this

action further in terms of conventional fields, we get [Flume *et. al.* (1996)]

$$\begin{aligned} \mathcal{A} = & \text{Tr} \int d^4x \left\{ \frac{1}{2} (\phi^\dagger D^2 \phi) + \bar{\psi} D \psi + g_o (\phi [\bar{\psi}, \gamma_5 \psi]) + g_o^2 [\phi^\dagger, \phi]^2 \right\} \\ & + \text{Tr} \int d^4x \left\{ \frac{1}{4g_o^2} \mathcal{F}^{\mu\nu} \mathcal{F}_{\mu\nu} + \frac{\theta_o}{32} \tilde{\mathcal{F}}^{\mu\nu} \mathcal{F}_{\mu\nu} \right\}. \end{aligned} \quad (4.208)$$

The action (4.208) can be recognized as the standard action for a Quark–Gluon–Higgs system in which all the fields are in the adjoint representation and the coupling constants are reduced to g and θ by the supersymmetry. Thus it is not very exotic. Indeed it could be the *QCD* action except for the fact that the quarks are in the adjoint and presence of the scalar field.

The action (4.208) embodies all the properties of *quantum gauge theory* that have surfaced over the past thirty years and could even be used as a model to teach quantum gauge theory. It might be worthwhile to list these properties [Flume *et. al.* (1996)]:

1. It contains a gauge-field coupled to matter
2. It is asymptotically free
3. It is scale-invariant, but with a scale-anomaly
4. It has spontaneous symmetry breaking
5. It has central charges (Z and \bar{Z})
6. It admits both instantons (see [Belavin *et. al.* (1975)]) and monopoles

Because of the supersymmetry it has some further special properties, whose significance will become clear later, namely,

7. It generalizes the *Montonen–Olive mass formula* [Montonen and Olive (1977)] for gauge-fields and monopoles:

$$\begin{aligned} \text{from} \quad M &= |v| \left(N_e + \frac{1}{g^2} n_m \right) & \text{to} \quad M &= |Z|, \\ \text{where} \quad Z &= (a n_e + a_d n_m), \end{aligned}$$

where n_e and n_m denote the gauge-field and monopole charges respectively, while the so-called *prepotential* coefficients a and a_d will be explained later.

8. It is symmetric with respect to a Z_4 symmetry, which is the relic of the R -symmetry: $(\theta_\alpha \rightarrow e^{i\epsilon} \theta_\alpha)$, which itself survives the axial anomaly breakdown.

9. It has a holomorphic structure
10. It has a duality that connects the weak and strong coupling regimes
12. The duality generalizes to an $SL(2, \mathbb{Z})$ symmetry.

4.14.1.3 Spontaneous Symmetry-Breaking

For $SU(2)$ this concept is very simple. From the form of the Higgs potential in (4.208) we see that there is a Higgs vacuum for $\phi = v\sigma$ where v is any complex number and σ is any fixed generator of $SU(2)$. Furthermore, for $v \neq 0$ this breaks the gauge-symmetry from $SU(2)$ to $U(1)$. For other gauge-groups G the corresponding statement is that \mathbf{v} must lie in the Cartan subalgebra of G . On the other hand there is no spontaneous breakdown of supersymmetry. Thus the full breakdown is

$$SU(2) \rightarrow U(1) : \quad N = 2 \quad \text{supersymmetry unbroken.}$$

Indeed it is the fact that the supersymmetry is unbroken that gives the model its nice properties, since otherwise the classical properties would not be preserved after quantization.

After the spontaneous breakdown the restriction of the $N = 1$ form of the classical action (4.207) to the massless $U(1)$ fields takes the form

$$\mathcal{A} = \text{Im} \int d^4x d^2\theta_\alpha d^2\bar{\theta}_\beta (\bar{A}A) + \tau_o \text{Im} \int d^4x d^2\theta_\alpha (W_\alpha W_\alpha).$$

Since the adjoint representation of $U(1)$ is trivial this action is a free-field one. However, in the quantum theory this does not mean that the effective Lagrangian is also free because, through the quantum fluctuations, the massive fields induce interaction term for the massless ones. The first great virtue of the SW model is that these interactions have a very specific form. In fact, they have shown that, due to the $N = 2$ supersymmetry the *local* part of the effective Lagrangian can only be of the form

$$\mathcal{A} = \frac{1}{2} \int d^4x d^2\theta d^2\bar{\theta} (\bar{A}A_d - \bar{A}_dA) + \text{Im} \int d^4x d^2\theta (\tau(A)) (W_\alpha W_\alpha), \quad (4.209)$$

where $A_d = \mathcal{F}'(A) \quad \text{and} \quad \tau(A) = \mathcal{F}''(A),$

for some function $\mathcal{F}(A)$. Thus the effective Lagrangian is completely governed by the single function $\mathcal{F}(A)$. Note that (4.209) is very similar to the classical action (4.207) which is the special case for which $\mathcal{F}(A) = \frac{1}{2}\tau_o A^2$.

The SW solution is actually a special Ansatz for the functional form of $\mathcal{F}(A)$ [Flume *et. al.* (1996)].

4.14.1.4 Holomorphy and Duality

It is now easy to quantify what is meant by holomorphy and duality. Holomorphy is simply the statement that $\mathcal{F}(A)$ depends only on A and not on \bar{A} . Duality means that the physics described by the effective action (4.209) is invariant with respect to the duality transformation [Flume *et. al.* (1996)]:

$$\begin{pmatrix} A \\ A_d \end{pmatrix} \rightarrow \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} A \\ A_d \end{pmatrix}, \quad D_\alpha W_\alpha \rightarrow D_{\dot{\alpha}} W_{\dot{\alpha}}, \quad \tau(A) \rightarrow (\tau(A))^{-1}. \quad (4.210)$$

Note that the duality transformation is closely linked to the Legendre transform of $\mathcal{F}(A)$ with respect to A . By noting that in the free classical theory with $\theta_o = 0$ the transformation (4.210) reduces to

$$\vec{E} \rightarrow \vec{B} \quad \text{and} \quad g \rightarrow \frac{1}{g},$$

we see that it is the generalization of the well-known *Maxwell–Dirac duality*. Thus the action (4.209) not only generalizes Maxwell–Dirac duality, but puts it into a genuine dynamical model. Furthermore, the duality (4.210) generalizes to

$$\begin{pmatrix} A \\ A_d \end{pmatrix} \rightarrow \begin{pmatrix} p & q \\ r & s \end{pmatrix} \begin{pmatrix} A \\ A_d \end{pmatrix} \quad \text{and} \quad \tau(A) \rightarrow \frac{p\tau(A) + q}{r\tau(A) + s},$$

where the matrix with entries (p, q, r, s) is in $SL(2, \mathbb{Z})$. The integer-valuedness of the transformation follows from the requirement that, in the perturbation theory at least, it should change the θ angle only by multiples of 2π and leaves the mass-formula form-invariant.

4.14.1.5 The SW Prepotential

The formulation of the SW solution [Seiberg and Witten (1994a); Seiberg and Witten (1994b)] itself is relatively simple, as follows (see, e.g., [Marshakov (1997)]). Supersymmetry requires the metric on *moduli space* of massless complex scalars from $N = 2$ vector supermultiplets to be of ‘special Kähler form’, i.e., the *Kähler potential*

$$K(a, \bar{a}) = \text{Im} \sum_i \bar{a}_i \frac{\partial \mathcal{F}}{\partial a_i}$$

needs to be expressed through a certain *holomorphic function* $\mathcal{F} = \mathcal{F}(a)$, called a *prepotential*. Then the Coulomb–branch, low–energy, effective action for the 4D $N = 2$ SUSY YM vector multiplets can be described in terms of an auxiliary *Riemann surface* (i.e., a complex curve) Σ , equipped with a *meromorphic 1–form* dS . The required Riemann surface Σ has some peculiar properties:

- The number of ‘live’ moduli (of complex structure) of Σ is strongly restricted (roughly ‘3 times’ less than for a generic Riemann surface). The genus of Σ for the $SU(N)$ gauge theories is exactly equal ²⁵ to the rank of gauge group – i.e., to the number of independent moduli.
- The variation of generating 1–form dS over these moduli gives *holomorphic* differentials.
- The periods of the generating 1–form

$$a = \oint_A dS \quad \text{and} \quad a_D = \oint_B dS$$

give the set of ‘dual’ masses, the W –bosons and the monopoles, while the period matrix $T_{ij}(\Sigma)$ represents the set of couplings in the low–energy effective theory. The *prepotential* $\mathcal{F} = \mathcal{F}(a)$ is a function of half of the variables (a, a_D) , so that we have

$$a_D^i = \frac{\partial \mathcal{F}}{\partial a_i} \quad \text{and} \quad T_{ij} = \frac{\partial a_D^i}{\partial a_j} = \frac{\partial^2 \mathcal{F}}{\partial a_i \partial a_j}.$$

These data mean that the effective SW theory is formulated in terms of a classical finite–gap integrable system (see, e.g., [Dubrovin *et. al.* (1985)]) and their *Whitham deformations* [Marshakov (1997)].

4.14.2 Clifford Actions, Dirac Operators and Spinor Bundles

Recall that the SW monopole equations were written in terms of a section of a *Spinor bundle* and a $U(1)$ connection on a line bundle L . The first

²⁵For generic gauge groups one should speak instead of *genus* (i.e., dimension of Jacobian of a spectral curve) – about the dimension of Prym–variety. Practically it means that for other than A_N –type gauge theories one should consider the spectral curves with *involution* and only the invariant under the involution cycles possess physical meaning. We consider in detail only the A_N theories, the generalization to the other gauge groups is straightforward: for example, instead of periodic Toda chains [Toda (1981)], corresponding to A_N theories, one has to consider the ‘generalized’ Toda chains, first introduced for different Lie–algebraic series (B, C, D, E, F and G) in [Bogoyavlensky (1981)].

equation just says that the spinor section ψ has to be in the kernel of the *Dirac operator*. The second equation describes a relation between the self-dual part of the curvature associated to the connection A and the section ψ in terms of the *Clifford action*.

The mathematical setting for Witten's gauge theory is considerably simpler than Donaldson's analogue: first of all it deals with $U(1)$ -principal bundles (Hermitian line bundles) rather than with $SU(2)$ -bundles, and the Abelian structure group allows simpler calculations; moreover the equation, which plays a role somehow analogous to the previous anti-self-dual equation for $SU(2)$ -instantons (see [Donaldson and Kronheimer (1990)]), involves Dirac operators and $Spin_c$ -structures, which are well known and long developed mathematical tools (see [Roe (1988)] or [Libermann and Marle (1987)]).

The main differences between the two theories arise when it comes to the properties of the moduli space of solutions of the monopole equation up to gauge transformations. The SW invariant, which depends on the Chern class of the line bundle L , is given by the number of points, counted with orientation, in a zero-dimensional moduli space.

In Witten's paper [Witten (1994)] the monopole equation is introduced, and the main properties of the moduli space of solutions are deduced.

The dimension of the moduli space is computed by an index theory technique, following an analogous proof for Donaldson's theory, as in [Atiyah *et al.* (1978)]; and the circumstances under which the SW invariants provide a topological invariant of the four-manifold are illustrated in a similar way to the analogous result regarding the Donaldson polynomials.

The tool that is of primary importance in proving the results about the moduli space of Abelian instantons is the Weitzenböck formula for the Dirac operator on the $Spin_c$ -bundle $S^+ \otimes L$: such a formula is a well known (see [Roe (1988)]) decomposition of the square of the Dirac operator on a spin bundle twisted with a line bundle L .

A first property which follows from the Weitzenböck formula is a bound on the number of solutions: the moduli space is empty for all but finitely many choices of the line bundle L .

Moreover, as shown in [Kronheimer and Mrowka (1994a)], the moduli space is always compact: a fact that avoids the complicated analytic techniques that were needed for the compactification of the moduli space of $SU(2)$ -instantons (see [Donaldson and Kronheimer (1990)]).

Another advantage of this theory is that the singularities of the moduli space (again this is shown in [Kronheimer and Mrowka (1994a)]) only ap-

pear at the trivial section $\psi \equiv 0$, since elsewhere the action of the gauge group is free. Hence, by perturbing the equation, it is possible to get a smooth moduli space.

The analogue of the vanishing Theorem for Donaldson polynomials on a manifold that splits as a connected sum can be proven, reinforcing the intuitive feeling that the two sets of invariants ought to be the same.

Moreover, explicit computations can be done in the case of Kähler manifolds, by looking at the SW invariants associated to the canonical line bundle.

The latter result has a generalization due to [Taubes (1994)], where it is shown that the value ± 1 of the SW invariants is achieved on symplectic four-manifolds, with respect to the canonical line bundle, by a technique that involves estimates of solutions of a parametrized family of perturbed monopole equations.

4.14.2.1 Clifford Algebras and Dirac Operators

Recall that the *Clifford algebra* $C(V)$ of a (real or complex) vector space V with a symmetric bilinear form $(,)$ is the algebra generated by the elements [Marcolli (1995)] $\{e_1^{\epsilon_1} \cdots e_n^{\epsilon_n},\}$, where $\epsilon_i = 0$, or 1 and $\{e_i\}$ is an orthogonal basis of V , subject to the relations $e \cdot e' + e' \cdot e = -2(e, e')$. The multiplication of elements of V in the Clifford algebra is called Clifford multiplication.

In particular given a differentiable manifold X we shall consider the Clifford algebra associated to the tangent space at each point. The Clifford algebra of the tangent bundle of X is the bundle that has fibre over each point $x \in X$ the Clifford algebra $C(T_x X)$. We shall denote this bundle $C(TX)$.

If $\dim V = 2m$, there is a unique irreducible representation of the Clifford algebra $C(V)$. This representation has dimension 2^m .

A *spinor bundle* over a Riemannian manifold X is the vector bundle associated to $C(TX)$ via this irreducible representation, endowed with a Hermitian structure such that the Clifford multiplication is skew-symmetric and compatible with the Levi-Civita connection on X (see [Roe (1988)]).

Not all manifolds admit a spinor bundle; it has been proved in [Roe (1988)] that the existence of such a bundle is equivalent to the existence of a $Spin_c$ -structure on the manifold X : we shall discuss $Spin_c$ -structures in the next paragraph. If such a bundle exists, it splits as a direct sum of two vector bundles, $S = S^+ \oplus S^-$, where the splitting is given by the internal grading of the Clifford algebra.

Let X be a manifold that admits a spinor bundle S . Let $\{e_i\}$ be a local orthonormal basis of sections of the tangent bundle TX and \langle, \rangle be the Hermitian structure as in the above definition of spinors. Then, for any section $\psi \in \Gamma(X, S)$, the expression $\langle e_i e_j \psi, \psi \rangle$ is purely imaginary at each point $x \in X$. As a proof, by skew-adjointness of Clifford multiplication and the fact that the basis is orthonormal,

$$\langle e_i e_j \psi, \psi \rangle = -\langle e_j \psi, e_i \psi \rangle = \langle \psi, e_j e_i \psi \rangle - \langle \psi, e_i e_j \psi \rangle = -\overline{\langle e_i e_j \psi, \psi \rangle}.$$

Given a spinor bundle S over X , the Dirac operator on S is a first order differential operator on the smooth sections $D : \Gamma(X, S^+) \rightarrow \Gamma(X, S^-)$, defined as the composition

$$D : \Gamma(X, S^+) \xrightarrow{\nabla} \Gamma(X, S^+) \otimes T^*X \xrightarrow{g} \Gamma(X, S^+) \otimes TX \xrightarrow{\bullet} \Gamma(X, S^-),$$

where the first map is the covariant derivative, with the *Spin-connection* induced by the *Levi-Civita connection* on X (see [Roe (1988)]), the second is the Legendre transform given by the Riemannian metric, and the third is Clifford multiplication.

It is easy to check (for details see [Roe (1988)]) that this corresponds to the following expression in coordinates: $Ds = e_k \cdot \nabla_k s$.

An essential tool in Spin geometry, which is very useful in SW gauge theory (see e.g. [Jost *et. al.* (1995); Kronheimer and Mrowka (1994a); Taubes (1994); Witten (1994)]), is the *Weitzenböck formula*.

Given a smooth vector bundle E over a *Spin_c*-manifold X and a connection A on E , the twisted Dirac operator $D_A : \Gamma(X, S^+ \otimes E) \rightarrow \Gamma(X, S^- \otimes E)$ is the operator acting on a section $s \otimes e$ as the Dirac operator on s and the composite of the covariant derivative $\tilde{\nabla}_A$ and the Clifford multiplication on e :

$$\begin{aligned} D_A : \Gamma(X, S^+ \otimes E) &\xrightarrow{\nabla \otimes 1 + 1 \otimes \tilde{\nabla}_A} \Gamma(X, S^+ \otimes E \otimes T^*X) \xrightarrow{g} \\ &\xrightarrow{g} \Gamma(X, S^+ \otimes E \otimes TX) \xrightarrow{\bullet} \Gamma(X, S^- \otimes E). \end{aligned}$$

The twisted Dirac operator D_A satisfies the *Weitzenböck formula*:

$$D_A^2 s = (\nabla_A^* \nabla_A + \frac{\kappa}{4} + \frac{-i}{4} F_A) s,$$

where ∇_A^* is the formal adjoint of the covariant derivative with respect to the *Spin-connection* on the Spinor bundle, and with respect to the connection A on E , $\nabla_A = \nabla \otimes 1 + 1 \otimes \tilde{\nabla}_A$; κ is the scalar curvature on

X , F_A is the curvature of the connection A , and $s \in \Gamma(X, S^+ \otimes E)$. As a proof, in local coordinates ($i \leq j$)

$$D_A^2 s = e_i \nabla_{A_i} (e_j \nabla_{A_j} s) = e_i e_j \nabla_{A_i} \nabla_{A_j} s = -\nabla_{A_i}^2 s + e_i e_j (\nabla_{A_i} \nabla_{A_j} - \nabla_{A_j} \nabla_{A_i}) s,$$

where $\nabla_A = \nabla \otimes 1 + 1 \otimes \tilde{\nabla}_A$. The first summand is $\nabla_A^* \nabla_A$ [Roe (1988)], and the second splits into a term which corresponds to the scalar curvature on X [Roe (1988)] and the curvature $-iF_A$ of the connection A (note that here we identify the Lie algebra of $U(1)$ with $i\mathbb{R}$).

4.14.2.2 *Spin and Spin_c Structures*

The group $Spin(n)$ is the universal covering of $SO(n)$.

The group $Spin_c(n)$ is defined via the following extension:

$$1 \rightarrow \mathbb{Z}_2 \rightarrow Spin_c(n) \rightarrow SO(n) \times U(1) \rightarrow 1, \quad (4.211)$$

i.e., $Spin_c(n) = (Spin(n) \times U(1))/\mathbb{Z}_2$.

The extension (4.211) determines the exact sheaf-cohomology sequence:

$$\cdots \rightarrow H^1(X; Spin_c(n)) \rightarrow H^1(X; SO(n)) \oplus H^1(X; U(1)) \xrightarrow{\delta} H^2(X; \mathbb{Z}_2). \quad (4.212)$$

By the standard fact that $H^1(X; G)$ represents the equivalence classes of principal G -bundles over X , we see that the connecting homomorphism of the sequence (4.212) is given by

$$\delta : (P_{SO(n)}, P_{U(1)}) \mapsto w_2(P_{SO(n)}) + \bar{c}_1(P_{U(1)}),$$

where $\bar{c}_1(P_{U(1)})$ is the reduction mod 2 of the first Chern class of the line bundle associated to the principal bundle $P_{U(1)}$ by the standard representation and w_2 is the second Stiefel-Whitney class.

A manifold X has a $Spin_c$ -structure if the frame bundle lifts to a principal $Spin_c(n)$ bundle. It has a $Spin$ -structure if it lifts to a $Spin(n)$ principal bundle.

From the above considerations on the cohomology sequence (4.212), and analogous considerations on the group $Spin(n)$, it follows that a manifold X admits a $Spin_c$ -structure iff $w_2(X)$ is the reduction mod 2 of an integral class. It has a $Spin$ -structure iff $w_2(X) = 0$. Different $Spin_c$ -structures on X are parametrized by $2H^2(X; \mathbb{Z}) \oplus H^1(X; \mathbb{Z}_2)$. This follows directly from (4.212): see [Libermann and Marle (1987)].

With this characterization of $Spin_c$ -structures we have the following Theorem, which has been proved in [Hirzebruch and Hopf (1958)]: Every oriented 4-manifold admits a $Spin_c$ -structure.

4.14.2.3 Spinor Bundles

Both $Spin(n)$ and $Spin_c(n)$ can be thought of as lying inside the Clifford algebra $C(\mathbb{R}^n)$, [Roe (1988); Libermann and Marle (1987)]. Therefore to a principal $Spin(n)$ or $Spin_c(n)$ bundle we can associate a vector bundle via the unique irreducible representation of the Clifford algebra. This will be the bundle of Spinors over X associated to the $Spin_c$ or $Spin$ structure, as defined in the definition of spinors above.

This can be given an explicit description in terms of transition functions (see [Libermann and Marle (1987)]). In fact let $g_{\alpha\beta}$ be the transition functions of the frame bundle over X , which take values in $SO(n)$. Then locally they can be lifted to functions $\tilde{g}_{\alpha\beta}$ which take values in $Spin(n)$, since on a differentiable manifold it is always possible to choose open sets with contractible intersections that trivialise the bundle.

However, if we have a $Spin_c$ -manifold that is not $Spin$, the $\tilde{g}_{\alpha\beta}$ will not form a cocycle, as $\tilde{g}_{\alpha\beta}\tilde{g}_{\beta\gamma}\tilde{g}_{\gamma\alpha} = 1$ means exactly that the second Stiefel-Whitney class vanishes.

Because of the $Spin_c$ structure we know that w_2 is the reduction of an integral class $c \in H^2(X, \mathbb{Z})$, which represents a complex line bundle, say with transition functions $\lambda_{\alpha\beta}$ with values in $U(1)$. Such functions will have a square root $\lambda_{\alpha\beta}^{1/2}$ locally; however, the line bundle will not have a square root globally (which is to say that the $\lambda_{\alpha\beta}^{1/2}$ won't form a cocycle), since by construction the first Chern class is not divisible by 2.

However, the relation $w_2(X) + c = 0 \bmod 2$ that comes from (4.212) says that the product $\tilde{g}_{\alpha\beta}\lambda_{\alpha\beta}^{1/2}$ is a cocycle. These are the transition functions of $S \otimes L$, where S would be the Spinor bundle of a $Spin$ structure and L would be the square root of a line bundle: neither of these objects is defined globally, but the tensor product is. This is the description of the Spinor Bundle of a $Spin_c$ structure that we shall use in the following.

Also, note that symplectic geometry plays a prominent role in the SW gauge theory. Many computations are possible for the case of symplectic 4-manifolds ([Taubes (1994)], [Taubes (1995a)], [Kotschik *et. al.* (1995)]); moreover, although the invariants are defined in terms of $U(1)$ -connections, and sections of Spinor bundles, as we shall illustrate below, they turn out to be strictly related to invariants of symplectic man-

ifolds, known as Gromov invariants (see [Gromov and Lawson (1980); Taubes (1995b)]). The fact that the SW invariants have a ‘more basic’ structure also led to a conjecture, suggested by Taubes, that symplectic manifolds may be among the most basic building blocks of the whole geometry of 4-manifolds.

4.14.2.4 The Gauge Group and Its Equations

Recall that the *gauge group* of a G –bundle is defined as the group of self equivalences of the bundle, namely the group of smooth maps

$$\lambda_\alpha : U_\alpha \rightarrow G, \quad \lambda_\beta = g_{\beta\alpha} \lambda_\alpha g_{\alpha\beta},$$

where the bundle is trivial over U_α and has transition functions $g_{\alpha\beta}$. It is often useful to consider this space of maps endowed with Sobolev norms, and, by completing with respect to the norm, to consider gauge groups of L_k^2 functions, as in [Donaldson and Kronheimer (1990); Freed and Uhlenbeck (1984)].

The gauge group is an infinite dimensional manifold; if the structure group is Abelian then it has a simpler description as $\mathcal{G} = \mathcal{M}(X, G)$, the space of maps from X to G .

In the case $G = U(1)$, the set of connected components of the gauge group \mathcal{G} is $H^1(X, \mathbb{Z})$.

The equations of the gauge theory are given in terms of a pair (A, ψ) of indeterminates, of which A is a connection on L and ψ is a smooth section of $S^+ \otimes L$.

The equations are

$$D_A \psi = 0, \quad (F_A^+)_{ij} = \frac{1}{4} \langle e_i e_j \psi, \psi \rangle e^i \wedge e^j,$$

where D_A is the Dirac operator twisted by the connection A , and F_A^+ is the self-dual part of the curvature associated to A . Here $\{e_i\}$ is a local basis of TX that acts on ψ by Clifford multiplication, $\{e^i\}$ is the dual basis of T^*X , and \langle, \rangle is the inner product on the fibres of $S^+ \otimes L$.

The gauge group of L is well defined although L is not globally defined as a line bundle, since the definition of the gauge group is given just in terms of the transition functions. In particular, as in the case of a line bundle, $\mathcal{G} = \mathcal{M}(X, U(1))$.

There’s an action of the gauge group on the space of pairs (A, ψ) , where

A is a connection on L and ψ a section of $S^+ \otimes L$, given by

$$\lambda : (A, \psi) \mapsto (A - 2i\lambda^{-1}d\lambda, i\lambda\psi). \quad (4.213)$$

The action defined in (4.213) induces an action of \mathcal{G} on the space of solutions to the SW equations. As a proof, it is enough to check that

$$D_{A-2i\lambda^{-1}d\lambda}(\lambda\psi) = i\lambda D_A\psi + i d\lambda \cdot \psi - 2i d\lambda \cdot \psi + i d\lambda \cdot \psi,$$

and in the second equation

$$F_{A-2i\lambda^{-1}d\lambda}^+ = F_A^+ - 2D^+(i\lambda^{-1}d\lambda) = F_A^+,$$

$$\text{and } \langle e_i e_j \lambda \psi, \lambda \psi \rangle = |\lambda|^2 \langle e_i e_j \psi, \psi \rangle = \langle e_i e_j \psi, \psi \rangle.$$

It is clear from (4.213) that the action of \mathcal{G} on the space of solutions is free iff ψ is not identically zero; while for $\psi \equiv 0$ the stabiliser of the action is $U(1)$, the group of constant gauge transformations.

4.14.3 Original SW Low Energy Effective Field Action

Now, let us examine the original SW theory [Seiberg and Witten (1994a); Seiberg and Witten (1994b)] in more detail. Recall that the classical potential of the pure $N = 2$ theory (without hypermultiplets) is

$$V(\phi) = \frac{1}{g^2} \text{Tr}[\phi, \phi^\dagger]^2. \quad (4.214)$$

For this to vanish, it is not necessary that ϕ should vanish; it is enough that ϕ and ϕ^\dagger commute. The classical theory therefore has a family of vacuum states. For instance, if the gauge group is $SU(2)$, then up to gauge transformation we can take $\phi = \frac{1}{2}a\sigma^3$, with $\sigma^3 = \text{diag}(1, -1)$ and a a complex parameter labelling the vacua. The *Weyl group* of $SU(2)$ acts by $a \leftrightarrow -a$, so the gauge invariant quantity parameterizing the space of vacua is $u = \frac{1}{2}a^2 = \text{Tr} \phi^2$. For non-zero a the gauge symmetry is broken to $U(1)$ and the global \mathbb{Z}_8 symmetry is broken to \mathbb{Z}_4 . The residual \mathbb{Z}_4 acts trivially on the u plane since the $U(1)_R$ charge of u is 4. The global symmetry group acts on the u plane as a spontaneously broken \mathbb{Z}_2 , acting by $u \leftrightarrow -u$. Classically, there is a singularity at $u = 0$, where the full $SU(2)$ gauge symmetry is restored and more fields become massless.

The next thing studied in [Seiberg and Witten (1994a); Seiberg and Witten (1994b)] was the low energy effective action of the light fields on the moduli space. For generic $\langle \phi \rangle$ the low energy effective Lagrangian contains a single $N = 2$ vector multiplet, \mathcal{A} . The terms with at most two derivatives

and not more than four fermions are constrained by $N = 2$ supersymmetry. They are expressed in terms of a single holomorphic function $\mathcal{F}(\mathcal{A})$. In $N = 1$ superspace, the Lagrangian is

$$\frac{1}{4\pi} \operatorname{Im} \left[\int d^4\theta \frac{\partial \mathcal{F}(A)}{\partial A} \bar{A} + \int d^2\theta \frac{1}{2} \frac{\partial^2 \mathcal{F}(A)}{\partial A^2} W_\alpha W^\alpha \right], \quad (4.215)$$

where A is the $N = 1$ chiral multiplet in the $N = 2$ vector multiplet \mathcal{A} whose scalar component is a . Here, Seiberg and Witten made the following comments:

1. For large a , asymptotic freedom takes over and the theory is weakly coupled. Moreover, since it is impossible to add an $N = 2$ invariant superpotential to (4.215), the vacuum degeneracy cannot be removed quantum mechanically. Therefore, the quantum theory has a non-trivial moduli space which is in fact a complex 1D *Kähler manifold*. The *Kähler potential* can be written in terms of the effective low energy \mathcal{F} function as

$$K = \operatorname{Im} \left(\frac{\partial \mathcal{F}(A)}{\partial A} \bar{A} \right).$$

The metric is thus concretely

$$(ds)^2 = \operatorname{Im} \left(\frac{\partial^2 \mathcal{F}(a)}{\partial a^2} da d\bar{a} \right). \quad (4.216)$$

In the classical theory, \mathcal{F} can be read off from the tree level Lagrangian of the $SU(2)$ gauge theory and is $\mathcal{F}(\mathcal{A}) = \frac{1}{2} \tau_{cl} A^2$ with $\tau_{cl} = \frac{\theta}{2\pi} + i \frac{4\pi}{g^2}$. Asymptotic freedom means that this formula is valid for large a if g^2 is replaced by a suitable effective coupling. The small a behavior will however turn out to be completely different. Classically, the θ parameter has no consequences. Quantum mechanically, the physics is θ dependent, but since there is an anomalous symmetry, it can be absorbed in a redefinition of the fields. Therefore, we can set $\theta = 0$.

2. The formula for the Kähler potential does not look covariant – the Kähler potential can be written in this way only in a distinguished class of coordinate systems. In fact, A is related by $N = 2$ supersymmetry to the ‘photon’ A_μ , which has a natural linear structure; this gives a natural coordinate system (or what will turn out to be a natural class of coordinate systems) for A .

3. The low energy values of the gauge coupling constant and theta parameter can be read off from the Lagrangian. If we combine them in

the form $\tau = \frac{\theta}{2\pi} + i\frac{4\pi}{g^2}$, and denote the effective couplings in the vacuum parametrized by a as $\tau(a)$, then $\tau(a) = \frac{\partial^2 \mathcal{F}}{\partial a^2}$.

4. The generalization to an arbitrary compact gauge group G of rank r is as follows. The potential is always given by (4.214), so the classical vacua are labelled by a complex adjoint-valued matrix ϕ with $[\phi, \phi^\dagger] = 0$. The unbroken gauge symmetry at the generic point on the moduli space is the *Cartan subalgebra* and therefore the complex dimension of the moduli space is r . The low energy theory is described in terms of r Abelian chiral multiplets \mathcal{A}^i , and the generalization of (4.215) is

$$\frac{1}{4\pi} \operatorname{Im} \left[\int d^4\theta \frac{\partial \mathcal{F}(A)}{\partial A^i} \bar{A}^i + \int d^2\theta \frac{1}{2} \frac{\partial^2 \mathcal{F}(A)}{\partial A^i \partial A^j} W_\alpha^i W^{\alpha j} \right]. \quad (4.217)$$

Here i labels the generators in the Cartan subalgebra and locally \mathcal{F} is an arbitrary holomorphic function of r complex variables.

5. The $SU(2)$ theory, studied on the flat direction with $u \neq 0$, has in addition to the massless chiral or vector multiplet \mathcal{A} , additional charged massive vector multiplets. One can easily write a gauge invariant effective action for the triplet of chiral multiplets \mathcal{A}^a , $a = 1 \dots 3$, which reduces at low energies to (4.215) for the massless fields and incorporates the massive ones. Using the same function \mathcal{F} as above, we set $\mathcal{F}(\sqrt{\mathcal{A} \cdot \mathcal{A}}) = \mathcal{H}(\mathcal{A} \cdot \mathcal{A})$ and write

$$\frac{1}{2\pi} \operatorname{Im} \left[\int d^4\theta \mathcal{H}' A^a (e^V)_{ab} \bar{A}^b + \int d^2\theta \frac{1}{2} \left(\mathcal{H}' \delta^{ab} + 2\mathcal{H}'' A^a A^b \right) W_\alpha^a W_\alpha^b \right], \quad (4.218)$$

where the $SU(2)$ -invariant metric δ^{ab} has been used to raise and lower indices. (4.218) has $N = 2$ supersymmetry and manifest gauge invariance, and reduces at low energies to (4.218).

6. The Lagrangian (4.218) is unchanged if we add to \mathcal{F} terms linear in \mathcal{A} . This has the effect of shifting $\partial \mathcal{F} / \partial \mathcal{A}$ by a constant.

As already mentioned, classically the \mathcal{F} function is

$$\mathcal{F}_0 = \frac{1}{2} \tau_{cl} \mathcal{A}^2. \quad (4.219)$$

The one-loop contributions add up to

$$\mathcal{F}_{one-loop} = i \frac{1}{2\pi} \mathcal{A}^2 \ln \frac{\mathcal{A}^2}{\Lambda^2}, \quad (4.220)$$

where Λ is the dynamically generated scale. This logarithm is related to the one-loop beta function and also ensures the anomalous transformation laws

under $U(1)_R$. Higher order perturbative corrections are absent. Instantons lead to new terms. The anomaly and the instanton action suggest that

$$\mathcal{F} = i \frac{1}{2\pi} \mathcal{A}^2 \ln \frac{\mathcal{A}^2}{\Lambda^2} + \sum_{k=1}^{\infty} \mathcal{F}_k \left(\frac{\Lambda}{\mathcal{A}} \right)^{4k} \mathcal{A}^2,$$

where the k 'th term arises as a contribution of k instantons. A detailed calculation of the $k = 1$ term indicates that $\mathcal{F}_1 \neq 0$. Also, corrections to the classical formula 4.219 are related to the beta function, and for $N = 4$ supersymmetric YM theory, whose beta function vanishes, the formula (4.219) is exact.

4.14.4 QED With Matter

Following the original SW approach [Seiberg and Witten (1994a); Seiberg and Witten (1994b)], we first consider Abelian gauge theories with $N = 2$ supersymmetry and charged matter hypermultiplets – that is, the $N = 2$ analog of ordinary QED.

The ‘photon’, A_μ is accompanied by its $N = 2$ superpartners – two neutral Weyl spinors λ and ψ that are often called ‘photinos’, and a complex neutral scalar a . They form an irreducible $N = 2$ representation that can be decomposed as a sum of two $N = 1$ representations: a and ψ are in a chiral representation, A , while A_μ and λ are in a vector representation, W_α .

We take the charged fields, the ‘electrons’, to consist of k hypermultiplets of electric charge one. Each hypermultiplet, for $i = 1 \dots k$, consists of two $N = 1$ chiral multiplets M^i and \tilde{M}_i with opposite electric charge; such an $N = 1$ chiral multiplet contains a Weyl fermion and a complex scalar.

The renormalizable $N = 2$ invariant Lagrangian is described in an $N = 1$ language by canonical kinetic terms and minimal gauge couplings for all the fields as well as a superpotential

$$W = \sqrt{2} A M^i \tilde{M}_i + \sum_i m_i M^i \tilde{M}_i. \quad (4.221)$$

The first term is related by $N = 2$ supersymmetry to the gauge coupling and the second one leads to $N = 2$ invariant mass terms.

The classical moduli space of the $N = 2$, $SU(2)$ gauge theory is parametrized by $u = \langle \text{Tr}(\phi^2) \rangle$ where ϕ is a complex scalar field in the adjoint representation of the gauge group. For $u \neq 0$ the gauge symmetry is broken to $U(1)$. At $u = 0$ the space is singular and the gauge symmetry

is unbroken. Our main goal is to determine – as quantitatively as possible – how this picture is modified quantum mechanically.

The quantum moduli space is described by the global supersymmetry version of special geometry. The *Kähler potential*, $K = \text{Im}(a_D(u)\bar{a}(\bar{u}))$, determines the metric (or, equivalently the kinetic terms). The pair (a_D, a) is a holomorphic section of an $SL(2, \mathbb{Z})$ bundle over the punctured complex u plane. They are related by $N = 2$ supersymmetry to a $U(1)$ gauge multiplet. a is related by $N = 2$ to the semiclassical ‘photon’ while a_D is related to its dual – ‘the magnetic photon’. The gauge kinetic energy is proportional to

$$\int d^2\theta \frac{\partial a_D}{\partial a} W_\alpha^2. \quad (4.222)$$

In this $N = 2$ theory, the one-loop approximation to K is exact (there are no higher order perturbative corrections and there are no $U(1)$ instantons on \mathbb{R}^4) leading to

$$a_D = -\frac{ik}{2\pi} a \log(a/\Lambda).$$

The lack of asymptotic freedom appears here as a breakdown of the theory at $|a| = \Lambda/e$, where the metric on the moduli space $\text{Im}(\frac{\partial a_D}{\partial a})$ vanishes and the effective gauge coupling is singular. This is the famous *Landau pole*.

For large $|u|$ the theory is semiclassical and

$$a \cong \sqrt{2u}, \quad a_D \cong i\frac{2}{\pi} a \log a. \quad (4.223)$$

These expressions are modified by instanton corrections. The exact expressions were determined as the periods on a torus

$$y^2 = (x^2 - \Lambda^4)(x - u) \quad (4.224)$$

of the *meromorphic 1-form* $\lambda = \frac{\sqrt{2}}{2\pi} \frac{dx}{y} \frac{(x-u)}{y}$. In (4.224), Λ is the dynamically generated mass scale of the theory.

The spectrum contains dyons labelled by various magnetic and electric charges. Stable states with magnetic and electric charges (n_m, n_e) have masses given by the *BPS formula*

$$M^2 = 2|Z|^2 = 2|n_e a(u) + n_m a_D(u)|^2. \quad (4.225)$$

There are two singular points on the quantum moduli space at $u = \pm\Lambda^2$; they are points at which a magnetic monopole becomes massless. When an

$N = 2$ breaking but $N = 1$ preserving mass term is added to the theory, these monopoles condense, leading to confinement (see below).

When the masses in (4.221) are not zero, the moduli space changes. The singularities on the Coulomb branch can move. Whenever $a = -\frac{1}{\sqrt{2}}m_i$ one of the electrons becomes massless. Therefore

$$a_D = -\frac{i}{2\pi} \sum_i (a + m_i/\sqrt{2}) \log \left(\frac{a + m_i/\sqrt{2}}{\Lambda} \right).$$

If some of the masses are equal, the corresponding singularities on the Coulomb branch coincide and there are more massless particles there. In this case a Higgs branch with non-zero expectation values for these electrons touches the Coulomb branch at the singularity. When there is only one massless electron hypermultiplet, the $|D|^2$ term in the potential prevents a Higgs branch from developing.

4.14.5 QCD With Matter

Following ([Seiberg and Witten (1994a); Seiberg and Witten (1994b)]), we now turn to QCD with an $SU(2)$ gauge group. The gluons are accompanied by Dirac fermions and complex scalars ϕ in the adjoint representation of the gauge group. We also add N_f hypermultiplets of quarks in the fundamental representation. As in the previous subsection, each hypermultiplet contains a Dirac fermion and four real scalars. In terms of $N = 1$ superfields the hypermultiplets contain two chiral superfields Q^i_a and \tilde{Q}_{ia} ($i = 1, \dots, N_f$ is the flavor index and $a = 1, 2$ the color index) and the $N = 2$ gauge multiplets include $N = 1$ gauge multiplets and chiral multiplets Φ . The superpotential for these chiral superfields is

$$W = \sqrt{2} \tilde{Q}_i \Phi Q^i + \sum_i m_i \tilde{Q}_i Q^i.$$

We now begin the analysis of the quantum moduli space. The first basic fact is that for large fields, the theory is weakly coupled and the quantum moduli space is well approximated by the classical moduli space. We parameterize the Coulomb branch by the gauge invariant coordinate $u = \langle \text{Tr } \phi^2 \rangle$.

For $N_f = 0$, the metric and the dyon masses are determined by a holomorphic section of an $SL(2, \mathbb{Z})$ bundle:

$$a = \frac{1}{2} \sqrt{2u} + \dots, \quad a_D = i \frac{4 - N_f}{2\pi} a(u) \log \frac{u}{\Lambda_{N_f}^2} + \dots,$$

where the ellipses represent instanton corrections and Λ_{N_f} is the dynamically generated scale of the theory with N_f flavors. The metric is $ds^2 = \text{Im}(a'_D \bar{a}') du d\bar{u}$ and the dyon masses $M^2 = 2|Z^2|$ are expressed in terms of $Z = n_e a + n_m a_D$, where n_m, n_e are the magnetic and electric charges, respectively.

4.14.6 Duality

Next, Seiberg and Witten performed $SL(2, \mathbb{Z})$ duality transformation on the low energy fields. Although they are non-local on the photon field A_μ , they act simply on (a_D, a) . Several new issues appeared when matter fields were present.

First, consider the situation of one massive quark with mass m_{N_f} and examine what happens when a approaches $m_{N_f}/\sqrt{2}$ where one of the elementary quarks becomes massless. Loop diagrams in which this quark propagates make a logarithmic contribution to a_D . The behavior near $a = m_{N_f}/\sqrt{2}$ is thus

$$a \approx a_0, \quad a_D \approx c - \frac{i}{2\pi}(a - a_0) \ln(a - a_0),$$

with $a_0 = m_{N_f}/\sqrt{2}$ and c a constant. The *monodromy*²⁶ around $a = a_0$ is

²⁶Recall that *monodromy* is the study of how geometrical objects behave as they ‘run around’ a singularity. It is closely associated with covering maps and their degeneration into ramification; the aspect giving rise to monodromy phenomena is that certain functions we may wish to define fail to be single-valued as we ‘run around’ a path encircling a singularity. The failure of monodromy is best measured by defining a *monodromy group*: a group of transformations acting on the data that codes what does happen as we ‘run around’.

In the case of a covering map, we look at monodromy as a special case of a fibration, and use the *homotopy lifting property* to ‘follow’ paths on the base space X (we assume it is path-connected, for simplicity) as they are lifted up into the cover C . If we follow round a loop based at a point $x \in X$, which we lift to start at c above x , we end at some c^* again above x ; it is quite possible that $c \neq c^*$, and to code this, one considers the action of the fundamental group $\pi_1(X, x)$ as a permutation group on the set of all c , as a monodromy group in this context.

An analogous geometrical role is played by parallel transport. In a principal bundle B over a smooth manifold M , a connection allows ‘horizontal’ movement from fibers above a point $m \in M$ to adjacent ones. The effect when applied to loops based at m is to define a *holonomy group* of translations of the fiber at m ; if the structure group of B is G , it is a subgroup of G that measures the deviation of B from the product bundle $M \times G$.

thus

$$a \rightarrow a, \quad a_D \rightarrow a_D + a - a_0 = a_D + a - \frac{m_{N_f}}{\sqrt{2}}. \quad (4.226)$$

Thus, under monodromy, the pair (a_D, a) is not simply transformed by $SL(2, \mathbb{Z})$; they also pick up additive constants. This possibility is not realized for the pure $N = 2$ gauge theory. The above simple consideration of a massless quark shows that this possibility does enter for $N_f > 0$.

If one arranges a_D , a , and the bare mass m as a 3D column vector $(m/\sqrt{2}, a_D, a)$, then the monodromy in (4.226) can be written in the general form

$$\mathcal{M} = \begin{pmatrix} 1 & 0 & 0 \\ r & k & l \\ q & n & p \end{pmatrix}, \quad (4.227)$$

with $\det \mathcal{M} = kp - nl = 1$. This is the most general form permitted by the low energy analysis. The specific form of the first row in (4.227) means that m is monodromy-invariant; intuitively this reflects the fact that m is a ‘constant’, not a ‘field’. Now, if one arranges the charges as a row vector $W = (S, n_m, n_e)$, then W transforms by $W \rightarrow W\mathcal{M}^{-1}$. Explicitly,

$$\mathcal{M}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ lq - pr & p & -l \\ nr - kq & -n & k \end{pmatrix}$$

Thus, the electric and magnetic charges n_e and n_m mix among themselves but do not get contributions proportional to the global symmetry charge S . On the other hand, the S charge can get contributions proportional to gauge charges n_e or n_m . Equivalently, the global symmetry can be transformed to a linear combination of itself and a gauge symmetry but not the other way around. Notice that the monodromy matrix mixing the charges in this way survives even if the bare mass m vanishes.

Now, it was noted above that locally, by virtue of $N = 2$ supersymmetry, the metric on the moduli space is of the form

$$(ds)^2 = \text{Im}(\tau(a) da d\bar{a}), \quad (4.228)$$

with $\tau(a)$ the holomorphic function $\tau = \partial^2 \mathcal{F} / \partial a^2$. The one-loop formula (4.220) shows that for large $|a|$, $\tau(a) \approx i (\ln(a^2/\Lambda^2) + 3) / \pi$ is a multivalued function whose imaginary part is single-valued and positive. However, if $\text{Im}[\tau(a)]$ is globally defined it cannot be positive definite as the harmonic

function $\text{Im}(\tau)$ cannot have a minimum. This indicates that the above description of the metric must be valid only locally.

To what extent is it possible to change variables from a to some other local parameter, while leaving the metric in the form (4.228)? The answer to this question is at the heart of the physics. We define $a_D = \partial\mathcal{F}/\partial a$. The metric can then be written

$$(ds)^2 = \text{Im}(da_D d\bar{a}) = -\frac{i}{2} (da_D d\bar{a} - da d\bar{a}_D). \quad (4.229)$$

This formula is completely symmetric in a and a_D , so if we use a_D as the local parameter, the metric will be in the same general form as (4.228), with a different harmonic function replacing $\text{Im}(\tau)$. This transformation corresponds to electric–magnetic duality.

4.14.6.1 Witten's Formalism

Now, following [Seiberg and Witten (1994a); Seiberg and Witten (1994b)], to be able to treat the formalism in a way that is completely symmetric between a and a_D , we introduce an arbitrary local holomorphic coordinate u , and treat a and a_D as functions of u , which is a local coordinate on a complex manifold \mathcal{M} – the moduli space of vacua of the theory. Eventually we pick u to be the expectation value of $\text{Tr}(\phi^2)$ – a good physical parameter – but for now u is arbitrary.

Introduce a 2D complex space $X \cong \mathbb{C}^2$ with coordinates (a_D, a) . Endow X with the symplectic form $\omega = \text{Im}(da_D \wedge d\bar{a})$. The functions $(a_D(u), a(u))$ give a map f from \mathcal{M} to X . The metric on \mathcal{M} is

$$(ds)^2 = \text{Im} \left(\frac{da_D}{du} \frac{d\bar{a}}{d\bar{u}} \right) du d\bar{u} = -\frac{i}{2} \left(\frac{da_D}{du} \frac{d\bar{a}}{d\bar{u}} - \frac{da}{du} \frac{d\bar{a}_D}{d\bar{u}} \right) du d\bar{u}. \quad (4.230)$$

This formula is valid for an arbitrary local parameter u on \mathcal{M} . If one picks $u = a$, one gets back the original formula (4.216) for the metric.²⁷ Notice that ω had no particular positivity property and thus, if $a(u)$ and $a_D(u)$ are completely arbitrary local holomorphic functions, the metric (4.230) is not positive. We will eventually construct $a(u)$ and $a_D(u)$ in a particular way that will ensure positivity.

It is easy to see what sort-of transformations preserve the general structure of the metric. If we set $a^\alpha = (a_D, a)$, $\alpha = 1, 2$, and let $\epsilon_{\alpha\beta}$ be the

²⁷This formula can be described in a coordinate-free way by saying that the *Kähler form* of the induced metric on \mathcal{M} is $f^*(\omega)$.

antisymmetric tensor with $\epsilon_{12} = 1$, then

$$(ds)^2 = -\frac{i}{2}\epsilon_{\alpha\beta}\frac{da^\alpha}{du}\frac{d\bar{a}^\beta}{d\bar{u}}dud\bar{u}. \quad (4.231)$$

This is manifestly invariant under linear transformations that preserve ϵ and commute with complex conjugation (the latter condition ensures that a^α and \bar{a}^α transform the same way). These transformations make the group $SL(2, \mathbb{R})$ (or equivalently $Sp(2, \mathbb{R})$). Also, (4.231) is obviously invariant under adding a constant to a_D or a . So if we arrange (a_D, a) as a column vector v , the symmetries that preserve the general structure are: $v \rightarrow Mv + c$, where M is a 2×2 matrix in $SL(2, \mathbb{R})$, and c is a constant vector. In general, this group of transformations can be thought of as the group of 3×3 matrices of the form $\begin{pmatrix} 1 & 0 \\ c & M \end{pmatrix}$, acting on the three objects $(1, a_D, a)$.

Generalization to Dimension Greater than One

Now, let us briefly discuss the generalization to other gauge groups. If the gauge group G has rank r , then \mathcal{M} has complex dimension r . Locally, from (4.217), it follows that the metric is

$$(ds)^2 = \text{Im} \left(\frac{\partial^2 \mathcal{F}}{\partial a^i \partial \bar{a}^j} \right) da^i d\bar{a}^j,$$

with distinguished local coordinates a^i and a holomorphic function \mathcal{F} . We again reformulate this by introducing

$$a_{D,j} = \frac{\partial \mathcal{F}}{\partial a^j}. \quad (4.232)$$

Then we can write

$$(ds)^2 = \text{Im} \sum_i da_{D,i} d\bar{a}^i.$$

To formulate this invariantly, we introduce a complex space $X \cong \mathbb{C}^{2r}$ with coordinates $a^i, a_{D,j}$. We endow X with the symplectic form

$$\omega = \frac{i}{2} \sum_i (da^i \wedge d\bar{a}_{D,i} - da_{D,i} \wedge d\bar{a}^i)$$

of type $(1, 1)$ and also with the holomorphic 2-form

$$\omega_h = \sum_i da^i \wedge da_{D,i}.$$

Then we introduce arbitrary local coordinates u^s , ($s = 1, \dots, r$), on the moduli space \mathcal{M} , and describe a map $f : \mathcal{M} \rightarrow X$ by functions $a^i(u)$, $a_{D,j}(u)$. We require f to be such that $f^*(\omega_h) = 0$; this precisely ensures that locally, if we pick $u^i = a^i$, then $a_{D,j}$ must be of the form in (4.232) with some holomorphic function \mathcal{F} . Then we take the metric on \mathcal{M} to be the one whose Kähler form is $f^*(\omega)$, i.e.,

$$(ds)^2 = \text{Im} \sum_{s,t,i} \frac{\partial a_{D,i}}{\partial u^s} \frac{\partial \bar{a}^i}{\partial \bar{u}^t} du^s d\bar{u}^t.$$

If again we arrange a, a_D as a $2r$ -component column vector v , then the formalism is invariant under transformations $v \rightarrow Mv + c$, with M a matrix in $Sp(2r, \mathbb{R})$ and c a constant vector. Again, considerations involving the charges will eventually require that M be in $Sp(2r, \mathbb{Z})$ and impose restrictions on c .

Physical Interpretation via Duality

So far we have seen that the spin zero component of the $N = 2$ multiplet has a *Kähler metric* of a very special sort, constructed using a distinguished set of coordinate systems. This rigid structure is related by $N = 2$ supersymmetry to the natural linear structure of the gauge field. We have found that, for the spin zero component, the distinguished parametrization is not completely unique; there is a natural family of parameterizations related by $SL(2, \mathbb{R})$. How does this $SL(2, \mathbb{R})$ (which will actually be reduced to $SL(2, \mathbb{Z})$) act on the gauge fields?

$SL(2, \mathbb{R})$ is generated by the transformations

$$T_b = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}, \quad \text{and} \quad S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

with real b . The former acts as $a_D \rightarrow a_D + ba$, $a \rightarrow a$; this acts trivially on the distinguished coordinate a , and can be taken to act trivially on the gauge field. By inspection of (4.218), the effect of $a_D \rightarrow a_D + ba$ on the gauge kinetic energy is just to shift the θ angle by $2\pi b$; in the Abelian theory, this has no effect until magnetic monopoles (or at least non-trivial $U(1)$ bundles) are considered. Once that is done, the allowed shifts in the θ angle are by integer multiples of 2π ; that is why b must be integral and gives essentially our first derivation of the reduction to $SL(2, \mathbb{Z})$.

The remaining challenge is to understand what S means in terms of the gauge fields. We will see that it corresponds to electric-magnetic duality. To

see this, let us see how duality works in Lagrangians of the sort introduced above.

We work in Minkowski space and consider first the purely bosonic terms involving only the gauge fields. We use conventions such that $\mathcal{F}_{\mu\nu}^2 = -(*\mathcal{F})_{\mu\nu}^2$ and $*(\mathcal{F}) = -\mathcal{F}$ where $*\mathcal{F}$ denotes the dual of \mathcal{F} . The relevant terms are

$$\frac{1}{32\pi} \operatorname{Im} \int \tau(a) \cdot (\mathcal{F} + i^* \mathcal{F})^2 = \frac{1}{16\pi} \operatorname{Im} \int \tau(a) \cdot (\mathcal{F}^2 + i^* \mathcal{F} \mathcal{F}). \quad (4.233)$$

Duality is carried out as follows. The constraint $d\mathcal{F} = 0$ (which in the original description follows from $\mathcal{F} = dA$) is implemented by adding a Lagrange multiplier vector field V_D . Then \mathcal{F} is treated as an independent field and integrated over. The normalization is set as follows. The $U(1) \subset SU(2)$ is normalized such that all $SO(3)$ fields have integer charges (matter multiplets in the fundamental representation of $SU(2)$ therefore have half integer charges). Then, a magnetic monopole corresponds to $\epsilon^{0\mu\nu\rho} \partial_\mu \mathcal{F}_{\nu\rho} = 8\pi\delta^{(3)}(x)$. For V_D to couple to it with charge one, we add to (4.233)

$$\frac{1}{8\pi} \int V_{D\mu} \epsilon^{\mu\nu\rho\sigma} \partial_\nu \mathcal{F}_{\rho\sigma} = \frac{1}{8\pi} \int * \mathcal{F}_D \mathcal{F} = \frac{1}{16\pi} \operatorname{Re} \int (*\mathcal{F}_D - i\mathcal{F}_D)(\mathcal{F} + i^* \mathcal{F}),$$

where $\mathcal{F}_{D\mu\nu} = \partial_\mu V_{D\nu} - \partial_\nu V_{D\mu}$ is the the field strength of V_D . We can now perform the Gaussian functional integral over \mathcal{F} and find an equivalent Lagrangian for V_D ,

$$\frac{1}{32\pi} \operatorname{Im} \left(\frac{-1}{\tau} \right) (\mathcal{F}_D + i^* \mathcal{F}_D)^2 = \frac{1}{16\pi} \operatorname{Im} \left(\frac{-1}{\tau} \right) (\mathcal{F}_D^2 + i^* \mathcal{F}_D \mathcal{F}_D).$$

We now repeat these steps in $N = 1$ superspace. We treat W_α in

$$\frac{1}{8\pi} \operatorname{Im} \int d^2\theta \tau(A) W^2$$

as an independent chiral field. The *superspace version of the Bianchi identity* $d\mathcal{F} = 0$ is

$$\operatorname{Im}(\mathcal{D}W) = 0,$$

where \mathcal{D} is the *supercovariant derivative*. It can be implemented by a real vector superfield V_D Lagrange multiplier. We add to the action

$$\frac{1}{4\pi} \operatorname{Im} \int d^4x d^4\theta V_D \mathcal{D}W = \frac{1}{4\pi} \operatorname{Re} \int d^4x d^4\theta i \mathcal{D}V_D W = -\frac{1}{4\pi} \operatorname{Im} \int d^4x d^2\theta W_D W.$$

Performing the Gaussian integral over W we find an equivalent Lagrangian

$$\frac{1}{8\pi} \operatorname{Im} \int d^2\theta \frac{-1}{\tau(A)} W_D^2. \quad (4.234)$$

To proceed further, we need to transform the $N = 1$ chiral multiplet A to A_D . The kinetic term

$$\operatorname{Im} \int d^4\theta h(A) \bar{A} \quad \text{is transformed by}$$

$$A_D = h(A) \quad \text{to} \quad \operatorname{Im} \int d^4\theta h_D(A_D) \bar{A}_D,$$

where $h_D(h(A)) = -A$ is minus the inverse function. Then using $h'(A) = \tau(A)$ the coefficient of the gauge kinetic term (4.234) becomes

$$-\frac{1}{\tau(A)} = -\frac{1}{h'(A)} = h'_D(A_D) = \tau_D(A_D).$$

Note that a shift of h by a constant does not affect the Lagrangian. Therefore, the duality transformation has a freedom to shift A_D by a constant.

The relations $A_D = h(A)$ and $h_D = -A$ mean that the duality transformation precisely implements the missing $SL(2, \mathbb{Z})$ generator S . The function $\tau = h'$ is mapped by $\tau_D(A_D) = -\frac{1}{\tau(A)}$. Remembering that $\tau(a) = \frac{\theta(a)}{2\pi} + i\frac{4\pi}{g(a)^2}$, we see that the duality transformation inverts τ rather than the low energy gauge coupling $g(a)$.

It is important to stress that unlike $\tau \rightarrow \tau + 1$, the duality transformation is not a symmetry of the theory. It maps one description of the theory to *another* description of the same theory.

For other gauge groups G the low energy Lagrangian has several Abelian fields, A^i , in the Cartan subalgebra.

Then $(A_D)_i = h_i(A^i) = \partial_i \mathcal{F}(A^i)$,

which leads to $h_D^i(h_j(A^k)) = -A^i$, and the ‘metrics’

$$\tau_{ij}(A) = \partial_i \partial_j \mathcal{F}(A) = \partial_j h_i(A), \quad \tau_D^{ij}(A_D) = \partial^i \partial^j \mathcal{F}_D(A_D) = \partial^j h_D^i(A_D)$$

satisfy $f_{ij} f_D^{jk} = -\delta_i^k$.

The above transformation together with the more obvious shifts: $A_{Di} \rightarrow A_{Di} + M_{ij} A^j$ – generate $Sp(2r, \mathbb{Z})$.

Coupling to Gravity

Now, we would like to compare the structure we have found to the ‘special geometry’ that appears if the chiral multiplet is coupled to $N = 2$ supergravity. In $N = 2$ supergravity, the general Kähler metric for a system of r chiral superfields is described locally by a holomorphic function $\mathcal{G}_0(a^1, \dots, a^r)$ of r complex variables a^i . The Kähler potential is

$$K_{grav} = -\ln \left(2i(\mathcal{G}_0 - \bar{\mathcal{G}}_0) + \frac{i}{2} \sum_i \left(\bar{a}^i \frac{\partial \mathcal{G}_0}{\partial a^i} - a^i \frac{\partial \bar{\mathcal{G}}_0}{\partial \bar{a}^i} \right) \right). \quad (4.235)$$

In global supersymmetry we had a local holomorphic function \mathcal{F} with

$$K = \frac{-i}{2} \sum_i \left(\bar{a}^i \frac{\partial \mathcal{F}}{\partial a^i} - a^i \frac{\partial \bar{\mathcal{F}}}{\partial \bar{a}^i} \right). \quad (4.236)$$

One would expect that there is some limit in which gravitational effects are small and (4.235) would reduce to (4.236). How does this occur?

It suffices to set

$$\mathcal{G}_0 = -i \frac{M_{Pl}^2}{4} + \mathcal{F}, \quad (4.237)$$

with M_{Pl} the Planck mass. Then if M_{Pl} is much larger than all relevant parameters, we get

$$K_{grav} = -\ln M_{Pl}^2 + \frac{K}{M_{Pl}^2} + O(M_{Pl}^{-4}).$$

The constant term $-\ln M_{Pl}^2$ does not contribute to the Kähler metric, so up to a normalization factor of $1/M_{Pl}^2$, the Kähler metric with supergravity reduces to that of global $N = 2$ supersymmetry as $M_{Pl} \rightarrow \infty$ keeping everything else fixed.

More fundamentally, we would like to compare the allowed monodromy groups. In supergravity, the global structure is exhibited as follows. One introduces an additional variable a^0 and sets $\mathcal{G} = (a^0)^2 \mathcal{G}_0$. One also introduces $a_{D,j} = \partial \mathcal{G} / \partial a^j$ for $j = 0, \dots, r$. Then one finds that the special Kähler structure of (4.235) allows $Sp(2r+2, \mathbb{R})$ transformations acting on (a_{Di}, a^j) .²⁸ Now, in decoupling gravity, we consider \mathcal{G} to be of the special form in (4.237). In that case, $a_{D,0} = -i \frac{M_{Pl}}{2}$. The other $a^i, a_{D,j}$ are independent of M_{Pl} . To preserve this situation in which M_{Pl} appears only in

²⁸In the gauge fields this is reduced to $Sp(2r+2, \mathbb{Z})$. The symplectic form preserved by $Sp(2r+2, \mathbb{R})$ becomes the usual one $\sum_i da^i \wedge da_{D,i}$.

$a_{D,0}$, we must consider only those $Sp(2r+2, \mathbb{R})$ transformations in which the transformations of all fields are independent of $a_{D,0}$. These transformations all leave a^0 invariant. There is no essential loss then in scaling the a 's so that $a^0 = 1$. Arrange the $a_{D,i}, a^j$ with $i, j = 1 \dots r$ as a column vector v . The $Sp(2r+2, \mathbb{R})$ transformations that leave invariant $a^0 = 1$ act on v by $v \rightarrow Mv + c$ where $M \in Sp(2r, \mathbb{R})$ and c is a constant. This is precisely the duality group that we found in the global $N = 2$ theory.

4.14.7 Structure of the Moduli Space

4.14.7.1 Singularity at Infinity

It is actually quite easy to see explicitly the appearance of non-trivial monodromies. In fact, asymptotic freedom implies a non-trivial monodromy at infinity. The renormalization group corrected classical formula $\mathcal{F}_{one\ loop} = i\mathcal{A}^2 \ln(\mathcal{A}^2/\Lambda^2)/2\pi$ gives for large a

$$a_D = \frac{\partial \mathcal{F}}{\partial a} \approx \frac{2ia}{\pi} \ln(a/\Lambda) + \frac{ia}{\pi}. \quad (4.238)$$

It follows that a_D is not a single-valued function of a for large a . If we recall that the physical parameter is really $u = \frac{1}{2}a^2$ (at least for large u and a), then the monodromy can be determined as follows. Under a circuit of the u plane at large u , one has $\ln u \rightarrow \ln u + 2\pi i$, and hence $\ln a \rightarrow \ln a + \pi i$. So the transformation is $a_D \rightarrow -a_D + 2a$, $a \rightarrow -a$. Thus, there is a non-trivial monodromy at infinity in the u plane,

$$M_\infty = PT^{-2} = \begin{pmatrix} -1 & 2 \\ 0 & -1 \end{pmatrix}, \quad (4.239)$$

where P is the element -1 of $SL(2, \mathbb{Z})$, and as usual $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

The factor of P in the monodromy exists already at the classical level. As we said above, a and $-a$ are related by a gauge transformation (the Weyl subgroup of the $SU(2)$ gauge group) and therefore we work on the u plane rather than its double cover, the a plane. In the anomaly free \mathbb{Z}_8 subgroup of the R symmetry group $U(1)_R$, there is an operation that acts on a by $a \rightarrow -a$; when combined with a Weyl transformation, this is the unbroken symmetry that we call P . Up to a gauge transformation it acts on the bosons by $\phi \rightarrow -\phi$, so it reverses the sign of the low energy electromagnetic field which in terms of $SU(2)$ variables is proportional to $\text{Tr}(\phi \mathcal{F})$. Hence it reverses the signs of all electric and magnetic charges and

acts as $-1 \in SL(2, \mathbb{Z})$. The P monodromy could be removed by (perhaps artificially) working on the a plane instead of the u plane.

The main new point here is the factor of T^{-2} which arises at the quantum level. This factor of T^{-2} has a simple physical explanation in terms of the electric charge of a magnetic monopole. Magnetic monopoles labelled by (n_m, n_e) have anomalous electric charge $n_e + \frac{\theta_{eff}}{2\pi} n_m$. The appropriate effective theta parameter is the low energy one

$$\theta_{eff} = 2\pi \operatorname{Re}[\tau(a)] = 2\pi \operatorname{Re} \left(\frac{da_D}{da} \right) = 2\pi \operatorname{Re} \left(\frac{da_D/du}{da/du} \right).$$

For large $|a|$, we have $\theta_{eff} \approx -4 \arg(a)$, which can be understood from the anomaly in the $U(1)_R$ symmetry. The monodromy at infinity transforms the row vector (n_m, n_e) to $(-n_m, -n_e - 2n_m)$, which implies that (a_D, a) transforms to $(-a_D + 2a, -a)$. The electric charge of the magnetic monopole can in fact be seen in the formula for Z , which if we take a_D from (4.238) and set $a = a_0 e^{-i\theta_{eff}/4}$ (with $a_0 > 0$) is

$$Z \approx a_0 e^{-i\theta_{eff}/4} \left\{ \left(n_e + \frac{\theta_{eff} n_m}{2\pi} \right) + i n_m \left(\frac{2 \ln a_0 / \Lambda + 1}{\pi} \right) \right\}.$$

The monodromy under $\theta_{eff} \rightarrow \theta_{eff} + 4\pi$ is easily seen from this formula to transform (n_m, n_e) in the expected fashion. Obviously, this simple formula depended on the semiclassical expression (4.238) for a_D ; with the exact expressions we presently propose, the results are much more complicated, in part because the effective theta angle is no longer simply the argument of a .

4.14.7.2 Singularities at Strong Coupling

The monodromy at infinity means that there must be an additional singularity somewhere in the u -plane. If \mathcal{M}' is the moduli space of vacua with all singularities deleted, then the monodromies must give a representation of the fundamental group of \mathcal{M}' in $SL(2, \mathbb{Z})$. Can this representation be Abelian? If the monodromies all commute with PT^{-2} , then a^2 is a good global complex coordinate, and the metric is globally of the form (4.228) with a global harmonic function $\operatorname{Im} \tau(a)$. As we have already noted, such a metric could not be positive.

The alternative is to assume a non-Abelian representation of the fundamental group. This requires at least two more punctures of the u plane (in addition to infinity). Since there is a symmetry $u \leftrightarrow -u$ acting on

the u plane, the minimal assumption is that there are precisely two more punctures exchanged by this symmetry.

The most natural physical interpretation of singularities in the u plane is that some additional massless particles are appearing at a particular value of u . For instance, in the classical theory, at $u = a = 0$, the $SU(2)$ gauge symmetry is restored; all the gluons become massless. In fact classically $a_D = 4\pi i a/g^2$ also vanishes at this point, and the monopoles and dyons become massless as well. One might be tempted to believe that the missing singularity comes from an analogous point in the quantum theory at which the gauge boson masses vanish. Though this behavior might seem unusual in asymptotically free theories in general, there are good indications that some $N = 1$ theories have an infrared fixed point with massless non-Abelian gluons.

4.14.7.3 Effects of a Massless Monopole

Now, following [Seiberg and Witten (1994a); Seiberg and Witten (1994b)], let us analyze the behavior of the effective Lagrangian near a point u_0 on the moduli space where magnetic monopoles become massless, that is, where

$$a_D(u_0) = 0.$$

Since monopoles couple in a non-local way to the original photon, we cannot use that photon in our effective Lagrangian. Instead, we should perform a duality transformation and write the effective Lagrangian in terms of the dual vector multiplet \mathcal{A}_D . The low energy theory is therefore an Abelian gauge theory with matter (an $N = 2$ version of QED). The unusual fact that the light matter fields are magnetically charged rather than electrically charged does not make any difference to the low energy physics. The only reason we call these particles monopoles rather than electrons is that this language is appropriate at large $|u|$ where the theory is semiclassical.

The dominant effect on the low energy gauge coupling constant is due to loops of light fields. In our case, these are the light monopoles. The low energy theory is not asymptotically free and therefore its gauge coupling constant becomes smaller as the mass of the monopoles becomes smaller. Since the mass is proportional to a_D , the low energy coupling goes to zero as $u \rightarrow u_0$. The electric coupling constant which is the inverse of the magnetic one diverges at that point.

More quantitatively, using the one-loop beta function, near the point where $a_D = 0$, the magnetic coupling is $\tau_D \approx -\frac{i}{\pi} \ln a_D$. Since a_D is a

good coordinate near that point, $a_D \approx c_0(u - u_0)$, with some constant c_0 . Using $\tau_D = dh_D/da_D$, we learn that

$$a(u) = -h_D(u) \approx a_0 + \frac{i}{\pi} a_D \ln a_D \approx a_0 + \frac{i}{\pi} c_0(u - u_0) \ln(u - u_0),$$

for some constant $a_0 = a(u = u_0)$. This constant a_0 cannot be zero because if it had been zero, all the electrically charged particles would have been massless at $u = u_0$ and the computation using light monopoles only would not be valid.

Now we can read off the monodromy. When u circles around u_0 ,

$$\begin{aligned} \text{so} \quad \ln(u - u_0) &\rightarrow \ln(u - u_0) + 2\pi i, \\ \text{then one has} \quad a_D &\rightarrow a_D, \quad a \rightarrow a - 2a_D. \end{aligned} \quad (4.240)$$

This effect is a sort-of dual of the monodromy at infinity. Near infinity, the monopole gains electric charge, and near $u = u_0$, the electron gains magnetic charge. (4.240) can be represented by the 2×2 monodromy matrix

$$M_1 = ST^2S^{-1} = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}. \quad (4.241)$$

4.14.7.4 The Third Singularity

With our assumption that there are only three singularities (counting $u = \infty$) and with two of the three monodromies determined in (4.239) and (4.241), we can now determine the third monodromy, which we call M_{-1} . With all of the monodromies taken in the counter clockwise direction, the monodromies must obey $M_1 M_{-1} = M_\infty$, and from this we get

$$M_{-1} = (TS)T^2(TS)^{-1} = \begin{pmatrix} -1 & 2 \\ -2 & 3 \end{pmatrix}.$$

The matrix M_{-1} is conjugate to M_1 . Actually,

$$\text{If} \quad A = TM_1 = \begin{pmatrix} -1 & 1 \\ -2 & 1 \end{pmatrix}, \quad (4.242)$$

$$\text{Then} \quad M_{-1} = AM_1A^{-1}. \quad (4.243)$$

Hence, M_{-1} can arise from a massless particle, just like M_1 . (4.243) would also hold if A is replaced by AM_1^r for any integer r .

What kind of particle should become massless to generate this singularity? If one arranges the charges as a row vector $q = (n_m, n_e)$, then

the massless particle that produces a monodromy M has $qM = q$. For instance, monodromy M_1 arises from a massless monopole of charge vector $q_1 = (1, 0)$, and using the known form of M_1 , one has $q_1 M_1 = q_1$. Duality symmetry implies that this must be so not just for the particular monodromy M_1 but for any monodromy coming from a massless particle. Upon setting $q_{-1} = (1, -1)$, we get $q_{-1} M_{-1} = q_{-1}$, and hence the monodromy M_{-1} arises from vanishing mass of a dyon of charges $(1, -1)$.

It seems that we are seeing massless particles of charges $(1, 0)$ or $(1, -1)$. However, there is in fact a complete democracy among dyons. The BPS-saturated dyons that exist semiclassically have charges $(1, n)$ (or $(-1, -n)$) for arbitrary integer n . The monodromy at infinity brings about a shift $(1, n) \rightarrow (1, n-2)$. If one carries out this shift n times before proceeding to the singularity at $u = 1$ or $u = -1$, the massless particles producing those singularities would have charges $(1, -2n)$ and $(1, -1-2n)$, respectively. This amounts to conjugating the representation of the fundamental group by M_∞^n .

The particular matrix A in (4.242) obeys $A^2 = -1$, which is equivalent to the identity as an automorphism of $SL(2, \mathbb{Z})$. Conjugation by A implements the underlying \mathbb{Z}_2 symmetry of the quantum moduli space, which according to our assumptions, exchanges the two singularities. The \mathbb{Z}_2 maps $M_1 \rightarrow M'_1 = M_{-1}$, $M_{-1} \rightarrow M'_{-1} = M_1$ and $M_\infty \rightarrow M'_\infty = M'_1 M'_{-1} = M_{-1} M_1$. Note that M'_∞ is not just obtained from M_∞ by conjugation, but the relation $M_\infty = M_1 M_{-1}$ is preserved. The reason for that is that (as in any situation in which one is considering a representation of the fundamental group of a manifold in a non-Abelian group), the definition of the monodromies requires a choice of base point. The operation $u \rightarrow -u$ acts on the base point, and this has to be taken into account in determining how M_∞ transforms under \mathbb{Z}_2 .

One can go farther and show that if one assumes the existence of a \mathbb{Z}_2 symmetry between M_1 and M_{-1} , then they must be conjugate to T^2 , and not some other power of T . In our derivation of the monodromy (4.241), the 2 came from something entirely independent of the assumption of a \mathbb{Z}_2 symmetry, namely, from the charges and multiplicities of the monopoles that exist semiclassically.

4.14.7.5 Monopole Condensation and Confinement

We recall that the underlying $N = 2$ chiral multiplet \mathcal{A} decomposes under $N = 1$ supersymmetry as a vector multiplet W_α and a chiral multiplet

Φ . Breaking $N = 2$ down to $N = 1$, one can add a superpotential $W = m\text{Tr}(\Phi^2)$ for the chiral multiplet. This gives a bare mass to Φ , reducing the theory at low energies to a pure $N = 1$ gauge theory. The low energy theory has a \mathbb{Z}_4 chiral symmetry. This theory is strongly believed to generate a mass gap, with confinement of charge and spontaneous breaking of \mathbb{Z}_4 to \mathbb{Z}_2 . Furthermore, there is no vacuum degeneracy except what is produced by this symmetry breaking, so that there are precisely two vacuum states.

How can this be mimicked in the low energy effective $N = 2$ theory? That theory has a moduli space \mathcal{M} of quantum vacua. The massless spectrum at least semiclassically consists solely of the Abelian chiral multiplet \mathcal{A} of the unbroken $U(1)$ subgroup of $SU(2)$. If those are indeed the only massless particles, the effect in the low energy theory of turning on m can be analyzed as follows. The operator $\text{Tr}(\Phi^2)$ is represented in the low energy theory by a chiral superfield U . Its first component is the scalar field u whose expectation value is

$$\langle u \rangle = \langle \text{Tr}(\phi^2) \rangle,$$

where ϕ is the $\theta = 0$ component of the superfield Φ . This is a holomorphic function on the moduli space. At least for small m we should add to our low energy Lagrangian an effective superpotential $W_{\text{eff}} = mU$.

Turning on the superpotential mU would perhaps eliminate almost all of the vacua and in the surviving vacua give a mass to the scalar components of \mathcal{A} . But if there are no extra degrees of freedom in the discussion, the gauge field in \mathcal{A} would remain massless. To get a mass for the gauge field, as is needed since the microscopic theory has a mass gap for $m \neq 0$, one needs either

- (i) extra light gauge fields, giving a non-Abelian gauge theory and possible strong coupling effects, or
- (ii) light charged fields, making possible a Higgs mechanism.

Thus we learn, as we did in discussing the monodromies, that somewhere on \mathcal{M} extra massless states must appear. The option (i) does not seem attractive, for reasons that we have already discussed. Instead we consider option (ii), with the further proviso, from our earlier discussion, that the light charged fields in question are monopoles and dyons.

Near the point at which there are massless monopoles, the monopoles can be represented in an $N = 1$ language by ordinary (local) chiral superfields M and \widetilde{M} , as long as we describe the gauge field by the dual to the

original photon, \mathcal{A}_D . The superpotential is

$$\widehat{W} = \sqrt{2}A_D M \widetilde{M} + mU(A_D), \quad (4.244)$$

where the first term is required by $N = 2$ invariance of the $m = 0$ theory, and the second term is the effective contribution to the superpotential induced by the microscopic perturbation $m \operatorname{Tr}(\Phi^2)$.

The presence of a term $m \operatorname{Tr}(\Phi^2)$ in the microscopic superpotential shows that the parameter m carries charge two. The low energy superpotential is holomorphic in its variables $\widehat{W}(m, M\widetilde{M}, A_D)$ and should have charge two under $U(1)_J$. Imposing that it is regular at $m = M\widetilde{M} = 0$, we find that it is of the form $\widehat{W} = m f_1(A_D) + M\widetilde{M} f_2(A_D)$. The functions f_1 and f_2 are independent of m and can be determined by examining the limit of small m , leading to (4.244).

The low energy vacuum structure is easy to analyze. Vacuum states correspond to solutions of $d\widehat{W} = 0$ (up to gauge transformation), which obey the additional condition $|M| = |\widetilde{M}|$ (we denote by M and \widetilde{M} both the superfields and their first components). The latter condition comes from vanishing of the D terms. Implementing these conditions, one finds if $m = 0$ that vacuum states correspond to $M = \widetilde{M} = 0$ with arbitrary a_D ; this is simply the familiar moduli space \mathcal{M} . If $m \neq 0$ the result is quite different. We get

$$\sqrt{2}M\widetilde{M} + m \frac{du}{da_D} = 0, \quad a_D M = a_D \widetilde{M} = 0.$$

Assuming that $du \neq 0$, the first equation requires $M, \widetilde{M} \neq 0$, while the second equation requires $a_D = 0$.

Expanding around this vacuum, it is easy to see that there is a mass gap. For instance, the gauge field gets a mass by the Higgs mechanism, since $M, \widetilde{M} \neq 0$. The Higgs mechanism in question is a magnetic Higgs mechanism, since the fields with expectation values are monopoles! Condensation of monopoles will induce confinement of electric charge. Thus, we get an explanation in terms of the low energy effective action of why the microscopic theory becomes confining when the $m \operatorname{Tr}(\Phi^2)$ superpotential is added.

4.14.8 *Masses and Periods*

The particle masses and the low energy metric and couplings were in [Seiberg and Witten (1994a); Seiberg and Witten (1994b)] determined by

equating a and a_D with periods of a certain meromorphic 1-form λ on the curve E . λ has two characteristics:

(i) λ may have poles but (as long as the monodromies are in $SL(2, \mathbb{Z})$) its residues vanish; and

(ii) To achieve positivity of the metric on the quantum moduli space, its derivative with respect to u is proportional to $\frac{dx}{y}$.

Condition (i) means that the definition of a and a_D by contour integrals

$$a = \oint_{\gamma_1} \lambda, \quad a_D = \oint_{\gamma_2} \lambda,$$

(with γ_1 and γ_2 some contours on E) is invariant under deformation of the γ_i , even across poles of λ . This ensures that only the homology classes of the γ_i matter and reduces the monodromies to a group $SL(2, \mathbb{Z})$ that acts on $H_1(E, \mathbb{Z})$. In the presence of bare masses, this is too strong a condition since when the bare masses are non-zero the monodromies are not quite in $SL(2, \mathbb{Z})$.

As for condition (ii), the differential form $\frac{dx}{y}$ has no poles and represents a cohomology class on E of type $(1, 0)$. Having $d\lambda/du = f(u) dx/y$ leads to positivity of the metric. The function $f(u)$ is determined by requiring the right behavior at the singularities, for instance $a \approx \frac{1}{2}\sqrt{2u}$ for large u , while f is a constant. The proper relation is in fact

$$\frac{d\lambda}{du} = \frac{\sqrt{2}}{8\pi} \frac{dx}{y}. \quad (4.245)$$

Up to an inessential sign, this is $1/2$ the value in the ‘old’ conventions. By integration with respect to u , (4.245) determines λ (once the curve is known) for all values of N_f . This relation is only supposed to hold up to a total differential in x ; λ is supposed to be meromorphic in x .

The massless $N_f = 3$ curve is given by

$$y^2 = x^2(x - u) - (x - u)^2.$$

The polynomial on the right hand side has zeroes at $x_0 = u$ and at

$$x_{\pm} = \frac{1}{2} (1 \pm \sqrt{1 - 4u}).$$

In particular, at $u = 1/4$, $x_+ = x_-$, giving the singularity that we have attributed to a massless state of $(n_m, n_e) = (2, 1)$. To show that this state is semiclassical, we interpolate on the positive u axis from the semiclassical

regime of $u \rightarrow \infty$ to the singularity at $u = 1/4$. For $u > 1/4$, x_+ and x_- are complex conjugates.

We have

$$\frac{da}{du} = \oint_{\gamma_1} \frac{d\lambda}{du} = \oint_{\gamma_1} \omega, \quad \frac{da_D}{du} = \oint_{\gamma_2} \frac{d\lambda}{du} = \oint_{\gamma_2} \omega,$$

where $\omega = (\sqrt{2}/8\pi)dx/y$, γ_1 is a circle in the x plane that loops around x_+ and x_- but not x_0 , and γ_2 is a contour that loops round x_0 and x_+ but not x_- . Complex conjugation leaves x_0 alone and exchanges x_+ with x_- ; hence γ_1 is invariant under complex conjugation, but complex conjugation turns γ_2 into a contour γ_3 that loops around x_0 and x_- while avoiding x_+ . So a is real but the complex conjugate of a_D is given by

$$\frac{d\bar{a}_D}{du} = \int_{\gamma_3} \omega. \quad (4.246)$$

γ_3 , however, is homotopic to the sum of $-\gamma_1$ and $-\gamma_2$ (the minus sign comes from keeping track of the orientations of the contours). Hence, (4.246) gives

$$\bar{a}_D = -a - a_D.$$

In other words,²⁹

$$a_D = -\frac{a}{2} + \text{imaginary}.$$

4.14.9 Residues

Since the jumps in a or a_D are integral linear combinations of $m_i/\sqrt{2}$ (with m_i the bare masses) and are $2\pi i$ times the residues of λ , the residues of λ should be of the form

$$\text{Res } \lambda = \sum_i \frac{n_i m_i}{2\pi i \sqrt{2}}, \quad \text{with } n_i \in \mathbb{Z}.$$

The $N_f = 4$ theory is controlled by a curve $y^2 = \mathcal{F}(x, u, m_i, \tau)$ and a

²⁹This equation has the following interpretation. The curve is real for real u , that is, the coefficients in the equation are real. There are two types of real elliptic curve: τ can have real part zero or $1/2$. (Thus τ is either invariant or transformed by the $SL(2, \mathbb{Z})$ transformation $\tau \rightarrow \tau - 1$ under the complex conjugation operation $\tau \rightarrow -\bar{\tau}$.) The two possibilities correspond in a suitable basis to a real, a_D imaginary, or a real, $a_D = -a/2 + \text{imaginary}$. For $u > 1/4$ we have the second possibility.

differential form λ obeying

$$\frac{d\lambda}{du} = \omega + \text{exact form in } x, \quad (4.247)$$

$$\text{with } \omega = \frac{\sqrt{2}}{8\pi} \frac{dx}{y}.$$

For $N = 4$ the structure is the same, except that 8π is replaced by 4π . \mathcal{F} should be such that the residues of λ are linear in the quark bare masses. This is a severe restriction on \mathcal{F} ; we see that it determines \mathcal{F} uniquely (up to the usual changes of variables) independently of most of the arguments that we have used up to this point.

Let us write (4.247) in a more symmetrical form. If $\lambda = dx a(x, u)$, then (4.247) means

$$\frac{\sqrt{2}}{8\pi} \frac{dx}{y} = dx \frac{\partial a}{\partial u} + dx \frac{\partial}{\partial x} f(x, u); \quad (4.248)$$

the arbitrary total x -derivative $dx \partial f / \partial x$ is allowed here because it does not contribute to the periods. (4.248) can be understood much better if written symmetrically in x and u . Henceforth, instead of using a 1-form $\omega = (\sqrt{2}/8\pi) \cdot dx/y$, we use a 2-form

$$\omega = \frac{\sqrt{2}}{8\pi} \frac{dx}{y} \frac{du}{y}.$$

Similarly, we combine the functions a, f appearing in (4.248) into a 1-form $\lambda = -a(x, u)dx + f(x, u)du$. The change in notation for ω and λ should cause no confusion. Then equation (4.248) can be more elegantly written as

$$\omega = d\lambda. \quad (4.249)$$

The meaning of the problem of finding λ can now be stated. Let X be the (noncompact) complex surface defined by the equation $y^2 = \mathcal{F}(x, u)$ (we suppress the parameters m_i and τ). Being closed, ω defines an element $[\omega] \in H^2(X, \mathbb{C})$. A smooth differential λ obeying (4.249) exists if and only if $[\omega] = 0$. Moreover, by standard theorems, in the absence of restrictions on the growth of λ at infinity, if λ exists it can be chosen to be holomorphic and of type $(1, 0)$.

If on the other hand $[\omega] \neq 0$, then (4.249) has no smooth, much less holomorphic, solution. However, X has the property that if one throws away a sufficient number of complex curves C_a , then $X' = X - \cup_a C_a$

has $H^2(X', \mathbb{C}) = 0$. (The necessary C_a are explicitly described later.) So if we restrict to X' , the cohomology class of ω vanishes and λ exists. λ may however have poles on the C_a , perhaps with residues, which we call $\text{Res}_{C_a}(\lambda)$.³⁰ If λ does have residues, then $d\lambda$ contains delta functions, and if one works on X instead of X' , one really has not (4.249) but

$$\omega = d\lambda - 2\pi i \sum_a \text{Res}_{C_a}(\lambda) \cdot [C_a] \quad (4.250)$$

where $[C_a]$ (which represents the cohomology class known as the Poincaré dual of C_a) is a delta function supported on C_a .

In cohomology, (4.250) simply means

$$[\omega] = -2\pi i \sum_a \text{Res}_{C_a}(\lambda) \cdot [C_a]. \quad (4.251)$$

Thus, if we pick the C_a so that the $[C_a]$ are a basis of $H^2(X, \mathbb{C})$, then the residues $\text{Res}_{C_a}(\lambda)$ are simply the coefficients of the expansion of $[\omega]$ in terms of the $[C_a]$. To find the residues we need not actually find λ ; it suffices to understand the cohomology class of ω by any method that may be available.

For instance, if X were compact, we could proceed as follows. First compute the intersection matrix

$$M_{ab} = \#(C_a \cdot C_b)$$

(that is, the number of intersection points of C_a and C_b , after perhaps perturbing the C_a so that they intersect generically). This is an invertible matrix. Second, calculate the periods

$$c_a = \int_{C_a} \omega.$$

Then

$$[\omega] = \sum_{a,b} c_a M_{ab}^{-1} [C_b].$$

Comparing to (4.246), we get

$$\text{Res}_{C_a}(\lambda) = -\frac{1}{2\pi i} \sum_b M_{ab}^{-1} c_b.$$

³⁰The residues of λ along C_a are constants, since $d\text{Res}_{C_a}(\lambda) = \text{Res}_{C_a} d\lambda = \text{Res}_{C_a} \omega = 0$.

4.14.10 *SW Monopole Equations and Donaldson Theory*

Developments in the understanding of $N = 2$ supersymmetric YM theory in 4D suggest a new point of view about *Donaldson theory* [Donaldson (1990); Donaldson (1986); Donaldson (1987); Donaldson and Kronheimer (1990)] of four manifolds: instead of defining 4-manifold invariants by counting $SU(2)$ instantons, one can define equivalent 4-manifold invariants by counting solutions of a nonlinear equation with an Abelian gauge group. This is a ‘dual’ equation in which the gauge group is the dual of the maximal torus of $SU(2)$. This new viewpoint, proposed by Witten in [Witten (1994)], suggests many new results about the Donaldson invariants.

Let X be an oriented, closed 4-manifold on which we pick a Riemannian structure with metric tensor g . $\Lambda^p T^*X$, or simply Λ^p , will denote the bundle of real-valued p -forms, and $\Lambda^{2,\pm}$ will be the sub-bundle of Λ^2 consisting of self-dual or anti-self-dual forms.

The monopole equations relevant to $SU(2)$ or $SO(3)$ Donaldson theory can be described as follows. If $w_2(X) = 0$, then X is a spin manifold and one can pick positive and negative spin bundles S^+ and S^- , of rank two.³¹ In that case, introduce a complex line bundle L ; the data in the monopole equation will be a connection A on L and a section M of $S^+ \otimes L$. The curvature 2-form of A will be called \mathcal{F} or $\mathcal{F}(A)$; its self-dual and anti-self-dual projections will be called \mathcal{F}^+ and \mathcal{F}^- .

If X is not spin, the S^\pm do not exist, but their projectivizations PS^\pm do exist (as bundles with fibers isomorphic to CP^1). A $Spin_c$ structure (which exists on any oriented four-manifold) can be described as a choice of a rank two complex vector bundle, which we write as $S^+ \otimes L$, whose projectivization is isomorphic to PS^+ . In this situation, L does not exist as a line bundle, but L^2 does;³² the motivation for writing the $Spin_c$ bundle as $S^+ \otimes L$ is that the tensor powers of this bundle obey isomorphisms suggested by the notation. For instance, $(S^+ \otimes L)^{\otimes 2} \cong L^2 \otimes (\Lambda^0 \oplus \Lambda^{2,+})$. The data of the monopole equation are now a section M of $S^+ \otimes L$ and a connection on $S^+ \otimes L$ that projects to the Riemannian connection on PS^+ . The symbol $\mathcal{F}(A)$ will now denote $1/2$ the trace of the curvature form of $S^+ \otimes L$.

Since L^2 is an ordinary line bundle, one has an integral cohomology class $x = -c_1(L^2) \in H^2(X, \mathbb{Z})$. Note that x reduces modulo two to $w_2(X)$;

³¹If there is more than one spin structure, the choice of a spin structure will not matter as we ultimately sum over twistings by line bundles.

³²One might be tempted to call this bundle L and write the $Spin_c$ bundle as $S^+ \otimes L^{1/2}$; that amounts to assigning magnetic charge $1/2$ to the monopole and seems unnatural physically.

in particular, if $w_2(X) = 0$, then L exists as a line bundle and $x = -2c_1(L)$.

To write the monopole equations, recall that S^+ is symplectic or pseudo-real, so that if M is a section of $S^+ \otimes L$, then the complex conjugate \bar{M} is a section of $S^+ \otimes L^{-1}$. The product $M \otimes \bar{M}$ would naturally lie in $(S^+ \otimes L) \otimes (S^+ \otimes L^{-1}) \cong \Lambda^0 \oplus \Lambda^{2,+}$. \mathcal{F}^+ also takes values in $\Lambda^{2,+}$ making it possible to write the following equations. Introduce Clifford matrices Γ_i (with anticommutators $\{\Gamma_i, \Gamma_j\} = 2g_{ij}$), and set $\Gamma_{ij} = \frac{1}{2}[\Gamma_i, \Gamma_j]$. Then the equations are³³

$$\mathcal{F}_{ij}^+ = -\frac{i}{2}\bar{M}\Gamma_{ij}M, \quad \sum_i \Gamma^i D_i M = 0.$$

In the second equation, $\sum_i \Gamma^i D_i$ is the *Dirac operator* D that maps sections of $S^+ \otimes L$ to sections of $S^- \otimes L$. We will sometimes abbreviate the first as $\mathcal{F}^+ = (M\bar{M})^+$. Alternatively, if positive spinor indices are written A, B, C , and negative spinor indices as A', B', C' ,³⁴ the equations can be written as

$$\mathcal{F}_{AB} = \frac{i}{2}(M_A \bar{M}_B + M_B \bar{M}_A), \quad D_{AA'} M^A = 0.$$

As a first step in understanding these equations, let us work out the virtual dimension of the moduli space \mathcal{M} of solutions of the equations up to gauge transformation. The linearization of the monopole equations fits into an elliptic complex

$$0 \rightarrow \Lambda^0 \xrightarrow{s} \Lambda^1 \oplus (S^+ \otimes L) \xrightarrow{t} \Lambda^{2,+} \oplus (S^- \otimes L) \rightarrow 0.$$

Here t is the linearization of the monopole equations, and s is the map from zero forms to deformations in A, M given by the infinitesimal action of the gauge group. Since we wish to work with real operators and determine the real dimension of \mathcal{M} , we temporarily think of $S^\pm \otimes L$ as real vector bundles (of rank four). Then an elliptic operator

$$T : \Lambda^1 \oplus (S^+ \otimes L) \rightarrow \Lambda^0 \oplus \Lambda^{2,+} \oplus (S^- \otimes L)$$

is defined by $T = s^* \oplus t$. The virtual dimension of the moduli space is given by the index of T . By dropping terms in T of order zero, T can be

³³To physicists the connection form A on a unitary line bundle is real; the covariant derivative is $d_A = d + iA$ and the curvature is $\mathcal{F} = dA$ or in components $\mathcal{F}_{ij} = \partial_i A_j - \partial_j A_i$.

³⁴Spinor indices are raised and lowered using the invariant tensor in $\Lambda^2 S^+$. In components, if $M^A = (M^1, M^2)$, then $M_A = (-M_2, M_1)$. One uses the same operation in interpreting \bar{M} as a section of $S^+ \otimes L$, so $\bar{M}^A = (M^2, -M^1)$. Also $\mathcal{F}_{AB} = \frac{1}{4}\mathcal{F}_{ij}\Gamma_{AB}^{ij}$.

deformed to the direct sum of the operator $d + d^*$ ³⁵ from Λ^1 to $\Lambda^0 \oplus \Lambda^{2,+}$ and the Dirac operator from $S^+ \otimes L$ to $S^- \otimes L$. The index of T is the index of $d + d^*$ plus twice what is usually called the index of the Dirac operator; the factor of two comes from looking at $S^\pm \otimes L$ as real bundles of twice the dimension. Let χ and σ be the Euler characteristic and signature of X . Then the index of $d + d^*$ is $-(\chi + \sigma)/2$, while twice the Dirac index is $-\sigma/4 + c_1(L)^2$. The virtual dimension of the moduli space is the sum of these or

$$W = -\frac{2\chi + 3\sigma}{4} + c_1(L)^2.$$

When this number is negative, there are generically no solutions of the monopole equations. When $W = 0$, that is, when $x = -c_1(L^2) = -2c_1(L)$ obeys

$$x^2 = 2\chi + 3\sigma, \quad (4.252)$$

then the virtual dimension is zero and the moduli space generically consists of a finite set of points $P_{i,x}$, $i = 1 \dots t_x$. With each such point, one can associate a sign $\epsilon_{i,x} = \pm 1$ – the sign of the determinant of T as we discuss momentarily. Once this is done, define for each x obeying (4.252) an integer n_x by

$$n_x = \sum_i \epsilon_{i,x}.$$

We will see later that $n_x = 0$ – indeed, the moduli space is empty – for all but finitely many x . Under certain conditions that we discuss in a moment, the n_x are topological invariants.

Note that $W = 0$ iff the index of the Dirac operator is

$$\Delta = \frac{\chi + \sigma}{4}.$$

In particular, Δ must be an integer to have non-trivial n_x . Similarly, Δ must be integral for the Donaldson invariants to be non-trivial (otherwise $SU(2)$ instanton moduli space is odd-dimensional).

For the sign of the determinant of T to make sense one must trivialize the determinant line of T . This can be done by deforming T as above to the direct sum of $d + d^*$ and the Dirac operator. If the Dirac operator, which naturally has a non-trivial *complex* determinant line, is regarded as a real operator, then its determinant line is naturally trivial – as a complex

³⁵What is meant here is a projection of the $d + d^*$ operator to self-dual forms.

line has a natural orientation. The $d + d^*$ operator is independent of A and M (as the gauge group is Abelian), and its determinant line is trivialized once and for all by picking an orientation of $H^1(X, \mathbb{R}) \oplus H^{2,+}(X, \mathbb{R})$. Note that this is the same data needed by Donaldson [Donaldson (1987)] to orient instanton moduli spaces for $SU(2)$; this is an aspect of the relation between the two theories.

If one replaces L by L^{-1} , A by $-A$, and M by \bar{M} , the monopole equations are invariant, but the trivialization of the determinant line is multiplied by $(-1)^\delta$ with δ the Dirac index. Hence the invariants for L and L^{-1} are related by

$$n_{-x} = (-1)^\Delta n_x.$$

For $W < 0$, the moduli space is generically empty. For $W > 0$ one can try, as in Donaldson theory, to define topological invariants that involve integration over the moduli space. Donaldson theory does not detect those invariants at least in known situations. We will see below that even when $W > 0$, the moduli space is empty for almost all x .

4.14.10.1 Topological Invariance

In general, the number of solutions of a system of equations weighted by the sign of the determinant of the operator analogous to T is always a topological invariant if a suitable compactness holds. If as in the case at hand one has a gauge invariant system of equations, and one wishes to count gauge orbits of solutions up to gauge transformations, then one requires (i) compactness; and (ii) free action of the gauge group on the space of solutions.

Compactness fails if a field or its derivatives can go to infinity. To explain the contrast with Donaldson theory, note that for $SU(2)$ instantons compactness fails precisely because an instanton can shrink to zero size. This is possible because

- (i) the equations are conformally invariant,
- (ii) they have non-trivial solutions on a flat \mathbb{R}^4 , and
- (iii) embedding such a solution, scaled to very small size, on any four-manifold gives a highly localized approximate solution of the instanton equations (which can sometimes be perturbed to an exact solution). The monopole equations by contrast are scale invariant but they have no non-constant L^2 solutions on flat \mathbb{R}^4 (or after dimensional reduction on flat \mathbb{R}^n with $1 \leq n \leq 3$). So there is no analog for the monopole equations of the

phenomenon where an instanton shrinks to zero size.

On the other hand, an obstruction does arise, just as in Donaldson theory from the question of whether the gauge group acts freely on the space of solutions of the monopole equations. The only way for the gauge group to fail to act freely is that there might be a solution with $M = 0$, in which case a constant gauge transformation acts trivially. A solution with $M = 0$ necessarily has $\mathcal{F}^+ = 0$, that is, it is an Abelian instanton.

Since $\mathcal{F}/2\pi$ represents the first *Chern class* of the line bundle L , it is integral; in particular if $\mathcal{F}^+ = 0$ then $\mathcal{F}/2\pi$ lies in the intersection of the integral lattice in $H^2(X, \mathbb{R})$ with the anti-self-dual subspace $H^{2,-}(X, \mathbb{R})$. As long as $b_2^+ \geq 1$, so that the self-dual part of $H^2(X, \mathbb{R})$ is non-empty, the intersection of the anti-self-dual part and the integral lattice generically consists only of the zero vector. In this case, for a generic metric on X , there are no Abelian instantons (except for $x = 0$, which we momentarily exclude) and n_x is well-defined.

To show that the n_x are topological invariants, one must further show that any two generic metrics on X can be joined by a path along which there is never an Abelian instanton. As in Donaldson theory, this can fail if $b_2^+ = 1$. In that case, the self-dual part of $H^2(X, \mathbb{R})$ is 1D, and in a generic one-parameter family of metrics on X , one may meet a metric for which there is an Abelian instanton. When this occurs, the n_x can jump. Let us analyze how this happens, assuming for simplicity that $b_1 = 0$. Given $b_1 = 0$ and $b_2^+ = 1$, one has $W = 0$ precisely if the index of the *Dirac equation* is 1. Therefore, there is generically a single solution M_0 of the Dirac equation $DM = 0$.

The equation $\mathcal{F}^+(A) = 0$ cannot be obeyed for a generic metric on X , but we want to look at the behavior near a special metric for which it does have a solution. Consider a one-parameter family of metrics parametrized by a real parameter ϵ , such that at $\epsilon = 0$ the self-dual subspace in $H^2(X, \mathbb{R})$ crosses a ‘wall’ and a solution A_0 of $\mathcal{F}^+(A) = 0$ appears. Hence for $\epsilon = 0$, we can solve the monopole equations with $A = A_0$, $M = 0$. Let us see what happens to this solution when ϵ is very small but non-zero. We set $M = mM_0$, with m a complex number, to obey $DM = 0$, and we write $A = A_0 + \epsilon\delta A$. The equation $\mathcal{F}^+(A) - (M\bar{M})^+ = 0$ becomes

$$\mathcal{F}^+(A_0) + (d\delta A)^+ - |m|^2(M_0\bar{M}_0)^+ = 0. \quad (4.253)$$

As the cokernel of $A \rightarrow \mathcal{F}^+(A)$ is 1D, δA can be chosen to project the left hand side of equation (4.253) into a 1D subspace (as $b_1 = 0$, this can be done in a unique way up to a gauge transformation). The remaining

equation looks near $\epsilon = 0$ like $c\epsilon - m\bar{m} = 0$, with c a constant. The ϵ term on the left comes from the fact that $\mathcal{F}^+(A_0)$ is proportional to ϵ .

Now we can see what happens for $\epsilon \neq 0$ to the solution that at $\epsilon = 0$ has $A = A_0$, $M = M_0$. Depending on the sign of c , there is a solution for m , uniquely determined up to gauge transformation, for $\epsilon > 0$ and no solution for $\epsilon < 0$, or vice-versa. Therefore n_x jumps by ± 1 , depending on the sign of c , in passing through $\epsilon = 0$.

The trivial Abelian instanton with $x = 0$ is an exception to the above discussion, since it cannot be removed by perturbing the metric. To define n_0 , perturb the equation $\mathcal{F}_{AB} = \frac{i}{2}(M_A\bar{M}_B + M_B\bar{M}_A)$ to

$$\mathcal{F}_{AB} = \frac{i}{2}(M_A\bar{M}_B + M_B\bar{M}_A) - p_{AB}, \quad (4.254)$$

with p a self-dual harmonic 2-form; with this perturbation, the gauge group acts freely on the solution space. Then define n_0 as the number of gauge orbits of solutions of the perturbed equations weighted by sign in the usual way. This is invariant under continuous deformations of p for $p \neq 0$; as long as $b_2^+ > 1$, so that the space of possible p 's is connected, the integer n_0 defined this way is a topological invariant.

The perturbation just pointed out will be used later in the case that p is the real part of a holomorphic 2-form to compute the invariants of Kähler manifolds with $b_2^+ > 1$. It probably has other applications; for instance, the case that p is a symplectic form is of interest.

4.14.10.2 Vanishing Theorems

Some of the main properties of the monopole equations can be understood by means of vanishing theorems. The general strategy in deriving such vanishing theorems is quite standard, but some unusual cancellations (required by the Lorentz invariance of the underlying untwisted theory) lead to unusually strong results.

If we set $s = \mathcal{F}^+ - M\bar{M}$, $k = DM$, then a small calculation gives

$$\begin{aligned} \int_X d^4x \sqrt{g} \left(\frac{1}{2}|s|^2 + |k|^2 \right) = \\ \int_X d^4x \sqrt{g} \left(\frac{1}{2}|\mathcal{F}^+|^2 + g^{ij}D_i M^A D_j \bar{M}_A + \frac{1}{2}|M|^4 + \frac{1}{4}R|M|^2 \right). \end{aligned} \quad (4.255)$$

Here g is the metric of X , R the scalar curvature, and $d^4x\sqrt{g}$ the Riemannian measure. A salient feature here is that a term $\mathcal{F}_{AB}M^A\bar{M}^B$, which

appears in either $|s|^2$ or $|k|^2$, cancels in the sum. This sharpens the implications of the formula, as we see. One can also consider the effect here of the perturbation in (4.254); the sole effect of this is to replace $\frac{1}{2}|M|^4$ in (4.255) by

$$\int_X d^4x \sqrt{g} \left(\mathcal{F}^+ \wedge p + \sum_{A,B} \left| \frac{1}{2} (M_A \bar{M}_B + M_B \bar{M}_A) - p_{AB} \right|^2 \right). \quad (4.256)$$

The second term is non-negative, and the first is simply the intersection pairing

$$2\pi c_1(L) \cdot [p]. \quad (4.257)$$

An obvious inference from (4.255) is that if X admits a metric whose scalar curvature is positive, then for such a metric any solution of the monopole equations must have $M = 0$ and $\mathcal{F}^+ = 0$. But if $b_2^+ > 0$, then after a generic small perturbation of the metric (which will preserve the fact that the scalar curvature is positive), there are no Abelian solutions of $\mathcal{F}^+ = 0$ except flat connections. Therefore, for such manifolds and metrics, a solution of the monopole equations is a flat connection with $M = 0$. These too can be eliminated using the perturbation in (4.254).³⁶ Hence a four-manifold for which $b_2^+ > 0$ and $n_x \neq 0$ for some x does not admit a metric of positive scalar curvature.

We can extend this to determine the possible four-manifolds X with $b_2^+ > 0$, some $n_x \neq 0$, and a metric of *non-negative* scalar curvature.³⁷ If X obeys those conditions, then for any metric of $R \geq 0$, any basic class x is in $H^{2,-}$ modulo torsion (so that L admits a connection with $\mathcal{F}^+ = 0$, enabling (4.255) to vanish); in particular if x is not torsion then $x^2 < 0$. Now consider the effect of the perturbation (4.254). As $x \in H^{2,-}$, (4.257) vanishes; hence if $R \geq 0$, R must be zero, M must be covariantly constant and $(M\bar{M})^+ = p$. For M covariantly constant, $(M\bar{M})^+ = p$ implies that p is covariantly constant also; but for all $p \in H^{2,+}$ to be covariantly constant implies that X is Kähler with $b_2^+ = 1$ or is hyper-Kähler. Hyper-Kähler metrics certainly have $R = 0$, and there are examples of metrics with $R = 0$ on Kähler manifolds with $b_2^+ = 1$ [LeBrun (1991)].

³⁶Flat connections can only arise if $c_1(L)$ is torsion; in that case, $c_1(L) \cdot [p] = 0$.

³⁷If $b_2^+ = 1$, the n_x are not all topological invariants, and we interpret the hypothesis to mean that with at least one sign of the perturbation in (4.254), the n_x are not all zero.

As an example, for a Kähler manifold with $b_2^+ \geq 3$, the canonical divisor K always arises as a basic class, so except in the hyper-Kähler case, such manifolds do not admit a metric of non-negative scalar curvature.

Even if the scalar curvature is not positive, we can get an explicit bound from (4.255) showing that there are only finitely many basic classes. Since

$$\int_X d^4x \sqrt{g} \left(\frac{1}{2} |M|^4 + \frac{1}{4} R |M|^2 \right) \geq -\frac{1}{32} \int_X d^4x \sqrt{g} R^2, \quad (4.258)$$

it follows from (4.255), even if we throw away the term $|D_i M|^2$, that

$$\int_X d^4x \sqrt{g} |\mathcal{F}^+|^2 \leq \frac{1}{16} \int_X d^4x \sqrt{g} R^2.$$

On the other hand, basic classes correspond to line bundles L with $c_1(L)^2 = (2\chi + 3\sigma)/4$, or

$$\frac{1}{(2\pi)^2} \int d^4x \sqrt{g} (|\mathcal{F}^+|^2 - |\mathcal{F}^-|^2) = \frac{2\chi + 3\sigma}{4}. \quad (4.259)$$

Therefore, for a basic class both $I^+ = \int d^4x \sqrt{g} |\mathcal{F}^+|^2$ and $I^- = \int d^4x \sqrt{g} |\mathcal{F}^-|^2$ are bounded. For a given metric, there are only finitely many isomorphism classes of line bundles admitting connections with given bounds on both I^+ and I^- , so the set of basic classes is finite.

The basic classes correspond to line bundles on which the moduli space of solutions of the monopole equations is of zero virtual dimension. We can analyze in a similar way components of the moduli space of positive dimension. Line bundles L such that $c_1(L)^2 < (2\chi + 3\sigma)/4$ are not of much interest in that connection, since for such line bundles the moduli space has negative virtual dimension and is generically empty. But if $c_1(L)^2 > (2\chi + 3\sigma)/4$, then (4.259) is simply replaced by the stronger bound

$$\frac{1}{(2\pi)^2} \int d^4x \sqrt{g} (|\mathcal{F}^+|^2 - |\mathcal{F}^-|^2) > \frac{2\chi + 3\sigma}{4}.$$

The set of isomorphism classes of line bundles admitting a connection obeying this inequality as well as (4.258) is once again finite. So we conclude that for any given metric on X , the set of isomorphism classes of line bundles for which the moduli space is non-empty and of non-negative virtual dimension is finite; for a generic metric on X , there are only finitely many non-empty components of the moduli space.

For further consequences of (4.255), we turn to a basic case in the study of four-manifolds: the case that X is Kähler.

4.14.10.3 *Computation on Kähler Manifolds*

If X is Kähler and spin, then $S^+ \otimes L$ has a decomposition $S^+ \otimes L \cong (K^{1/2} \otimes L) \oplus (K^{-1/2} \otimes L)$, where K is the canonical bundle and $K^{1/2}$ is a square root. If X is Kähler but not spin, then $S^+ \otimes L$, defined as before, has a similar decomposition where now $K^{1/2}$ and L are not defined separately and $K^{1/2} \otimes L$ is characterized as a square root of the line bundle $K \otimes L^2$.

We denote the components of M in $K^{1/2} \otimes L$ and in $K^{-1/2} \otimes L$ as α and $-i\bar{\beta}$, respectively. The equation $\mathcal{F}^+(A) = M\bar{M}$ can now be decomposed as

$$\mathcal{F}^{2,0} = \alpha\beta, \quad \mathcal{F}_{\omega}^{1,1} = -\frac{\omega}{2} (|\alpha|^2 - |\beta|^2), \quad \mathcal{F}^{0,2} = \bar{\alpha}\bar{\beta}. \quad (4.260)$$

Here ω is the Kähler form and $\mathcal{F}_{\omega}^{1,1}$ is the $(1,1)$ part of \mathcal{F}^+ . (4.255) can be rewritten

$$\begin{aligned} \int_X d^4x \sqrt{g} \left(\frac{1}{2} |s|^2 + |k|^2 \right) &= \int_X d^4x \sqrt{g} \left(\frac{1}{2} |\mathcal{F}^+|^2 + g^{ij} D_i \bar{\alpha} D_j \alpha + g^{ij} D_i \bar{\beta} D_j \beta \right. \\ &\quad \left. + \frac{1}{2} (|\alpha|^2 + |\beta|^2)^2 + \frac{1}{4} R (|\alpha|^2 + |\beta|^2) \right). \end{aligned} \quad (4.261)$$

The right hand side of (4.261) is not manifestly non-negative (unless $R \geq 0$), but the fact that it is equal to the left hand side shows that it is non-negative and zero precisely for solutions of the monopole equations. Consider the operation

$$A \rightarrow A, \quad \alpha \rightarrow \alpha, \quad \beta \rightarrow -\beta. \quad (4.262)$$

This is not a symmetry of the monopole equations. But it is a symmetry of the right hand side of (4.261). Therefore, given a zero of the right hand side of (4.261) – that is, a solution of the monopole equations – the operation (4.262) gives another zero of the right hand side of (4.261) – that is, another solution of the monopole equations. So, though not a symmetry of the monopole equations, the transformation (4.262) maps solutions of those equations to other solutions.

Vanishing of $\alpha\beta$ means that $\alpha = 0$ or $\beta = 0$. If $\alpha = 0$, then the Dirac equation for M reduces to

$$\bar{\partial}_A \beta = 0,$$

where $\bar{\partial}_A$ is the $\bar{\partial}$ operator on L . Similarly, if $\beta = 0$, then the Dirac equation gives

$$\bar{\partial}_A \alpha = 0.$$

Knowing that either α or β is zero, we can deduce which it is. Integrating the $(1, 1)$ part of (4.260) gives

$$\frac{1}{2\pi} \int_X \omega \wedge \mathcal{F} = -\frac{1}{4\pi} \int_X \omega \wedge \omega (|\alpha|^2 - |\beta|^2). \quad (4.263)$$

The left hand side of (4.263) is a topological invariant which can be interpreted as

$$J = [\omega] \cdot c_1(L).$$

The condition that there are no non-trivial Abelian instantons is that J is non-zero; we only wish to consider metrics for which this is so. If $J < 0$, we must have $\alpha \neq 0$, $\beta = 0$, and if $J > 0$, we must have $\alpha = 0$, $\beta \neq 0$.

The equation that we have not considered so far is the $(1, 1)$ part of (4.260). This equation can be interpreted as follows. Suppose for example that we are in the situation with $\beta = 0$. The space of connections A and sections α of $K^{1/2} \otimes L$ can be interpreted as a symplectic manifold, the symplectic structure being defined by

$$\langle \delta_1 A, \delta_2 A \rangle = \int_X \omega \wedge \delta_1 A \wedge \delta_2 A, \quad \langle \delta_1 \alpha, \delta_2 \alpha \rangle = -i \int_X \omega \wedge \omega (\delta_1 \bar{\alpha} \delta_2 \alpha - \delta_2 \bar{\alpha} \delta_1 \alpha).$$

On this symplectic manifold acts the group of $U(1)$ gauge transformations. The moment map μ for this action is the quantity that appears in the $(1, 1)$ equation that we have not yet exploited, that is

$$\mu\omega = \mathcal{F}_\omega^{1,1} + \omega|\alpha|^2.$$

By analogy with many similar problems, setting to zero the moment map and dividing by the group of $U(1)$ gauge transformations should be equivalent to dividing by the complexification of the group of gauge transformations. In the present case, the complexification of the group of gauge transformations acts by $\alpha \rightarrow t\alpha$, $\bar{\partial}_A \rightarrow t\bar{\partial}_A t^{-1}$, where t is a map from X to \mathbf{C}^* .

Conjugation by t has the effect of identifying any two A 's that define the same complex structure on L . This can be done almost uniquely: the ambiguity is that conjugation by a constant t does not change A . Obviously, a gauge transformation by a constant t rescales α by a constant. The result therefore, for $J < 0$, is that the moduli space of solutions of the monopole equations is the moduli space of pairs consisting of a complex structure on L and a non-zero holomorphic section, defined up to scaling, of $K^{1/2} \otimes L$.

For $J > 0$, it is instead β that is non-zero, and $K^{1/2} \otimes L$ is replaced by $K^{1/2} \otimes L^{-1}$.

This result can be stated particularly nicely if X has $b_1 = 0$. Then the complex structure on L , assuming that it exists, is unique. The moduli space of solutions of the monopole equations is therefore simply a complex projective space, $PH^0(X, K^{1/2} \otimes L)$ or $PH^0(X, K^{1/2} \otimes L^{-1})$, depending on the sign of J .

Identifying The Basic Classes

We would now like to identify the basic classes. The above description of the moduli space gives considerable information: basic classes are of the form $x = -2c_1(L)$, where L is such that $J < 0$ and $H^0(X, K^{1/2} \otimes L)$ is non-empty, or $J > 0$ and $H^0(X, K^{1/2} \otimes L^{-1})$ is non-empty. This, however, is not a sharp result.

That is closely related to the fact that the moduli spaces $PH^0(X, K^{1/2} \otimes L^{\pm 1})$ found above very frequently have a dimension bigger than the ‘generic’ dimension of the moduli space as predicted by the index Theorem. In fact, Kähler metrics are far from being generic. In case the expected dimension is zero, one would have always $n_x > 0$ if the moduli spaces behaved ‘generically’ (given the complex orientation, an isolated point on the moduli space would always contribute $+1$ to n_x ; this is a special case of a discussion below). Since the n_x are frequently negative, moduli spaces of non-generic dimension must appear.

When the moduli space has greater than the generically expected dimension, one can proceed by integrating over the bosonic and fermionic collective coordinates in the path integral. This gives a result that can be described topologically: letting T be the operator that arises in linearizing the monopole equations, the cokernel of T is a vector bundle V (the ‘bundle of antighost zero modes’) over the moduli space \mathcal{M} ; its Euler class integrated over \mathcal{M} is the desired n_x .

Alternatively, one can perturb the equations to more generic ones. We use the same perturbation as before. For a Kähler manifold X , the condition $b_2^+ > 1$ is equivalent to $H^{2,0}(X) \neq 0$, so we can pick a non-zero holomorphic 2-form η . We perturb the monopole equations (4.260) to

$$\mathcal{F}^{2,0} = \alpha\beta - \eta, \quad \mathcal{F}_\omega^{1,1} = -\omega(|\alpha|^2 - |\beta|^2), \quad \mathcal{F}^{0,2} = \bar{\alpha}\bar{\beta} - \bar{\eta}, \quad (4.264)$$

leaving unchanged the Dirac equation for M .

It suffices to consider the case that the first Chern class of L is of type $(1, 1)$, since the unperturbed moduli space vanishes otherwise. That being so, we have

$$0 = \int_X F^{2,0} \wedge \bar{\eta} = \int_X F^{0,2} \wedge \eta.$$

Using this, one finds that (4.261) generalizes to

$$\begin{aligned} \int_X d^4x \sqrt{g} \left(\frac{1}{2} |s|^2 + |k|^2 \right) &= \int_X d^4x \left(\frac{1}{2} |\mathcal{F}^+|^2 + g^{ij} D_i \bar{\alpha} D_j \alpha + g^{ij} D_i \bar{\beta} D_j \beta \right. \\ &\quad \left. + \frac{1}{2} (|\alpha|^2 - |\beta|^2)^2 + 2|\alpha\beta - \eta|^2 + \frac{R}{4} (|\alpha|^2 + |\beta|^2) \right). \end{aligned} \quad (4.265)$$

Witten now makes an argument of a sort we have already seen: the transformation

$$A \rightarrow A, \quad \alpha \rightarrow \alpha, \quad \beta \rightarrow -\beta, \quad \eta \rightarrow -\eta, \quad (4.266)$$

though not a symmetry of (4.264), is a symmetry of the right hand side of (4.265). As solutions of (4.264) are the same as zeroes of the right hand side of (4.265), we deduce that the solutions of (4.264) with a 2-form η are transformed by (4.266) to the solutions with $-\eta$. The terms in (4.264) even or odd under the transformation must therefore separately vanish, so any solution of (4.264) has

$$0 = \mathcal{F}^{0,2} = \mathcal{F}^{2,0} = \alpha\beta - \eta.$$

The condition $\mathcal{F}^{0,2} = 0$ means that the connection still defines a holomorphic structure on L .

The condition $\alpha\beta = \eta$ gives our final criterion for determining the basic classes: they are of the form $x = -2c_1(L)$ where, for any choice of $\eta \in H^0(X, K)$, one has a factorization $\eta = \alpha\beta$ with holomorphic sections α and β of $K^{1/2} \otimes L^{\pm 1}$, and $x^2 = c_1(K)^2$.

To make this completely explicit, suppose the divisor of η is a union of irreducible components C_i of multiplicity r_i . Thus the canonical divisor is

$$c_1(K) = \sum_i r_i [C_i],$$

where $[C_i]$ denotes the cohomology class that is Poincaré dual to the curve C_i . The existence of the factorization $\eta = \alpha\beta$ means that the divisor of

$K^{1/2} \otimes L$ is

$$c_1(K^{1/2} \otimes L) = \sum_i s_i [C_i],$$

where s_i are integers with $0 \leq s_i \leq r_i$. The first Chern class of L is therefore

$$c_1(L) = \sum_i (s_i - \frac{1}{2}r_i)[C_i].$$

And the basic classes are of the form $x = -2c_1(L)$ or

$$x = - \sum_i (2s_i - r_i)[C_i].$$

An x of this form is of type $(1, 1)$ and congruent to $c_1(K)$ modulo two, but may not obey $x^2 = c_1(K)^2$. It is actually possible to prove using the Hodge index Theorem that for x as above, $x^2 \leq c_1(K)^2$. This is clear from the monopole equations: perturbed to $\eta \neq 0$, these equations have at most isolated solutions (from the isolated factorization $\eta = \alpha\beta$) and not a moduli space of solutions of positive dimension. So, for Kähler manifolds, the non-empty perturbed moduli spaces are at most of dimension zero; invariants associated with monopole moduli spaces of higher dimension vanish.

4.14.11 *SW Theory and Integrable Systems*

Remarkably, the SW theory for $N = 2$ supersymmetric YM theory for arbitrary gauge algebra \mathfrak{g} appears to be intimately related with the existence of certain classical mechanics integrable systems. This relation was first suspected on the basis of the similarity between the SW curves and the spectral curves of certain integrable models [Gorskii *et. al.* (1995); Nakatsu and Takasaki (1996)]. Then, arguments were developed that SW theory naturally produces integrable structures [Donagi and Witten (1996)].

For the $N = 2$ supersymmetric YM theory with massive hypermultiplet, the relevant integrable system appears to be the *elliptic Calogero–Moser system* [Calogero (1975); Moser (1975)]. For $SU(N)$ gauge group, Donagi and Witten [Donagi and Witten (1996)] proposed that the spectral curves of the $SU(N)$ Hitchin system should play the role of the SW curves. [Nekrasov (1996)] recognized that the $SU(N)$ Hitchin system spectral curves are identical to those of the $SU(N)$ elliptic Calogero–Moser (CM) integrable system. That the $SU(N)$ elliptic CM-curves (and associated SW differential) do indeed provide the SW solution for the

$N = 2$ theory with one massive hypermultiplet was fully established by the authors in [D'Hoker and Phong (1998a); D'Hoker and Phong (1998b); D'Hoker and Phong (1998c)], where it was shown that:

(i) The resulting effective prepotential \mathcal{F} (and thus the low energy effective action) reproduces correctly the logarithmic singularities predicted by perturbation theory.

(ii) \mathcal{F} satisfies a renormalization group type equation which determines explicitly and efficiently instanton contributions to any order.

(iii) The prepotential in the limit of large hypermultiplet mass m (as well as large gauge scalar expectation value and small gauge coupling) correctly reproduces the prepotentials for $N = 2$ super-YM theory with any number of hypermultiplets in the fundamental representation of the gauge group.

The $N = 2$ theory for arbitrary gauge algebra \mathfrak{g} and with one massive hypermultiplet in the adjoint representation was one such outstanding case when $\mathfrak{g} \neq SU(N)$. Actually, as discussed previously, upon taking suitable limits, this theory contains a very large number of models with smaller hypermultiplet representations R , and in this sense has a universal aspect. It appeared difficult to generalize directly the Donagi–Witten construction of Hitchin systems to arbitrary \mathfrak{g} , and it was thus natural to seek this generalization directly amongst the elliptic CM integrable systems. It has been known now for a long time, thanks to the work of Olshanetsky and Perelomov [Olshanetsky and Perelomov (1976); Olshanetsky and Perelomov (1981)], that CM-systems can be defined for any simple Lie algebra. Olshanetsky and Perelomov also showed that the CM-systems for *classical Lie algebras* were integrable, although the existence of a spectral curve (or, a Lax pair with a spectral parameter) as well as the case of exceptional Lie algebras remained open. Thus several immediate questions are:

(i) Does the elliptic CM-system for general Lie algebra \mathfrak{g} admit a Lax pair with spectral parameter?

(ii) Does it correspond to the $N = 2$ supersymmetric gauge theory with gauge algebra \mathfrak{g} and a hypermultiplet in the adjoint representation?

(ii) Can this correspondence be verified in the limiting cases when the mass m tends to 0 with the theory acquiring $N = 4$ supersymmetry and when $m \rightarrow \infty$, with the hypermultiplet decoupling in part to smaller representations of \mathfrak{g} ?

According to [D'Hoker and Phong (1998a); D'Hoker and Phong (1998b); D'Hoker and Phong (1998c)], the answers to these questions can be stated succinctly as follows:

(i) The elliptic CM-systems defined by an arbitrary simple Lie algebra \mathfrak{g} do admit Lax pairs with spectral parameters.

(ii) The correspondence between elliptic \mathfrak{g} CM-systems and $N = 2$ supersymmetric \mathfrak{g} gauge theories with matter in the adjoint representation holds directly when the Lie algebra \mathfrak{g} is simply-laced. When \mathfrak{g} is not simply-laced, the correspondence is with new integrable models, *the twisted elliptic CM-systems*.

(iii) The new twisted elliptic CM-systems also admit a Lax pair with spectral parameter.

(iv) In the scaling limit $m = Mq^{-\frac{1}{2}\delta} \rightarrow \infty$, (with M fixed), the twisted (respectively untwisted) elliptic \mathfrak{g} CM-systems tend to the Toda system for $(\mathfrak{g}^{(1)})^\vee$ (respectively $\mathfrak{g}^{(1)}$) for $\delta = \frac{1}{h_g^\vee}$ (respectively $\delta = \frac{1}{h_g}$). Here h_g and h_g^\vee are the Coxeter and the dual Coxeter numbers of \mathfrak{g} .

4.14.11.1 $SU(N)$ Elliptic CM System

The original elliptic CM-system is the system defined by the Hamiltonian

$$H(x, p) = \frac{1}{2} \sum_{i=1}^N p_i^2 - \frac{1}{2} m^2 \sum_{i \neq j} \wp(x_i - x_j). \quad (4.267)$$

Here m is a mass parameter, and $\wp(z)$ is the Weierstrass \wp -function, defined on a torus $\mathbb{C}/(2\omega_1\mathbb{Z} + 2\omega_2\mathbb{Z})$. As usual, we denote by $\tau = \omega_2/\omega_1$ the moduli of the torus, and set $q = e^{2\pi i\tau}$. The well-known trigonometric and rational limits with respective potentials

$$-\frac{1}{2} m^2 \sum_{i \neq j} \frac{1}{4 \operatorname{sh}^2\left(\frac{x_i - x_j}{2}\right)} \quad \text{and} \quad -\frac{1}{2} m^2 \sum_{i \neq j} \frac{1}{(x_i - x_j)^2},$$

arise in the limits $\omega_1 = -i\pi$, $\omega_2 \rightarrow \infty$ and $\omega_1, \omega_2 \rightarrow \infty$. All these systems have been shown to be completely *integrable in the sense of Liouville*, i.e., they all admit a complete set of integrals of motion which are in involution.

However, we require a notion of integrability which is in some sense more stringent, namely the existence of a Lax pair $L(z)$, $M(z)$ with spectral parameter z . The Hamiltonian system (4.267) is equivalent to the *Lax equation*

$$\dot{L}(z) = [L(z), M(z)], \quad (4.268)$$

with $L(z)$ and $M(z)$ given by the following $N \times N$ matrices

$$\begin{aligned} L_{ij}(z) &= p_i \delta_{ij} - m(1 - \delta_{ij})\Phi(x_i - x_j, z), \\ M_{ij}(z) &= m\delta_{ij} \sum_{k \neq i} \wp(x_i - x_k) - m(1 - \delta_{ij})\Phi'(x_i - x_j, z). \end{aligned} \quad (4.269)$$

The function $\Phi(x, z)$ is defined by

$$\Phi(x, z) = \frac{\sigma(z - x)}{\sigma(z)\sigma(x)} e^{x\zeta(z)}, \quad (4.270)$$

where $\sigma(z)$, $\zeta(z)$ are the usual Weierstrass σ and ζ functions on the torus $\mathbb{C}/(2\omega_1\mathbb{Z} + 2\omega_2\mathbb{Z})$. The function $\Phi(x, z)$ satisfies the key functional equation

$$\Phi(x, z)\Phi'(y, z) - \Phi(y, z)\Phi'(x, z) = (\wp(x) - \wp(y))\Phi(x + y, z). \quad (4.271)$$

It is well-known that functional equations of this form are required for the Hamilton equations of motion to be equivalent to the Lax equation (4.268) with a Lax pair of the form (4.269). Often, solutions had been obtained under additional parity assumptions in x (and y), which prevent the existence of a spectral parameter. The solution $\Phi(x, z)$ with spectral parameter z is obtained by dropping such parity assumptions for general z . Conversely, general functional equations of the form (4.271) essentially determine $\Phi(x, z)$.

4.14.11.2 CM Systems Defined by Lie Algebras

The Hamiltonian system (4.267) is only one example of a whole series of Hamiltonian systems associated with each simple Lie algebra. More precisely, given any simple Lie algebra \mathfrak{g} , we have the system with Hamiltonian [Olshanetsky and Perelomov (1976); Olshanetsky and Perelomov (1976)]

$$H(x, p) = \frac{1}{2} \sum_{i=1}^r p_i^2 - \frac{1}{2} \sum_{\alpha \in \mathcal{R}(\mathfrak{g})} m_{|\alpha|}^2 \wp(\alpha \cdot x), \quad (4.272)$$

where r is the rank of \mathfrak{g} , $\mathcal{R}(\mathfrak{g})$ denotes the set of roots of \mathfrak{g} , and the $m_{|\alpha|}$ are mass parameters. To preserve the invariance of the Hamiltonian (4.272) under the Weyl group, the parameters $m_{|\alpha|}$ depend only on the orbit $|\alpha|$ of the root α , and not on the root α itself. In the case of $A_{N-1} = SU(N)$, it is common practice to use N pairs of dynamical variables (x_i, p_i) , since the roots of A_{N-1} lie conveniently on a hyperplane in \mathbb{C}^N . The dynamics of the system are unaffected if we shift all x_i by a constant, and the number of degrees of freedom is effectively $N - 1 = r$. Now the roots of $SU(N)$ are given

by $\alpha = e_i - e_j$, $1 \leq i, j \leq N$, $i \neq j$. Thus we recognize the original elliptic CM-system as the special case of (4.272) corresponding to A_{N-1} . As in the original case, the elliptic systems (4.272) admit rational and trigonometric limits. Olshanetsky and Perelomov [Olshanetsky and Perelomov (1976); Olshanetsky and Perelomov (1976)] succeeded in constructing a Lax pair for all these systems in the case of classical Lie algebras, albeit without spectral parameter.

4.14.11.3 Twisted CM-Systems Defined by Lie Algebras

It turns out that the Hamiltonian systems (4.272) are not the only natural extensions of the basic elliptic CM-system. A subtlety arises for simple Lie algebras \mathfrak{g} which are not simply-laced, i.e., algebras which admit roots of uneven length. This is the case for the algebras B_n , C_n , G_2 , and F_4 in Cartan's classification. For these algebras, the following *twisted* elliptic CM-systems were introduced by the authors in [D'Hoker and Phong (1998a); D'Hoker and Phong (1998b); D'Hoker and Phong (1998c)]

$$H_{\mathfrak{g}}^{\text{twisted}} = \frac{1}{2} \sum_{i=1}^r p_i^2 - \frac{1}{2} \sum_{\alpha \in \mathcal{R}(\mathfrak{g})} m_{|\alpha|}^2 \wp_{\nu(\alpha)}(\alpha \cdot x). \quad (4.273)$$

Here the function $\nu(\alpha)$ depends only on the length of the root α . If \mathfrak{g} is simply-laced, we set $\nu(\alpha) = 1$ identically. Otherwise, for \mathfrak{g} non simply-laced, we set $\nu(\alpha) = 1$ when α is a long root, $\nu(\alpha) = 2$ when α is a short root and \mathfrak{g} is one of the algebras B_n , C_n , or F_4 , and $\nu(\alpha) = 3$ when α is a short root and $\mathfrak{g} = G_2$. The *twisted* Weierstrass function $\wp_{\nu}(z)$ is defined by

$$\wp_{\nu}(z) = \sum_{\sigma=0}^{\nu-1} \wp\left(z + 2\omega_a \frac{\sigma}{\nu}\right), \quad (4.274)$$

where ω_a is any of the half-periods ω_1 , ω_2 , or $\omega_1 + \omega_2$. Thus the twisted and untwisted CM-systems coincide for \mathfrak{g} simply laced. The original motivation for twisted CM-systems was based on their scaling limits.

4.14.11.4 Scaling Limits of CM-Systems

For the standard elliptic CM-systems corresponding to A_{N-1} , in the scaling limit we have [Inozemtsev (1989a); Inozemtsev (1989b)]

$$m = Mq^{-\frac{1}{2N}}, \quad q \rightarrow 0, \quad (4.275)$$

$$x_i = X_i - 2\omega_2 \frac{i}{N}, \quad 1 \leq i \leq N, \quad (4.276)$$

where M is kept fixed, the elliptic A_{N-1} CM Hamiltonian tends to the following Hamiltonian

$$H_{\text{Toda}} = \frac{1}{2} \sum_{i=1}^N p_i^2 - \frac{1}{2} \left(\sum_{i=1}^{N-1} e^{X_{i+1}-X_i} + e^{X_1-X_N} \right). \quad (4.277)$$

The roots $e_i - e_{i+1}$, $1 \leq i \leq N-1$, and $e_N - e_1$ can be recognized as the simple roots of the affine algebra $A_{N-1}^{(1)}$. Thus (4.277) can be recognized as the Hamiltonian of the Toda system defined by $A_{N-1}^{(1)}$.

Scaling Limits based on the Coxeter Number

The key feature of the above scaling limit is the collapse of the sum over the entire root lattice of A_{N-1} in the CM Hamiltonian to the sum over only simple roots in the Toda Hamiltonian for the Kac-Moody algebra $A_{N-1}^{(1)}$. Our task is to extend this mechanism to general Lie algebras. For this, we consider the following generalization of the preceding scaling limit

$$m = Mq^{-\frac{1}{2}\delta}, \quad x = X - 2\omega_2\delta\rho^\vee, \quad (4.278)$$

Here $x = (x_i)$, $X = (X_i)$ and ρ^\vee are r D vectors. The vector x is the dynamical variable of the CM-system. The parameters δ and ρ^\vee depend on the algebra \mathfrak{g} and are yet to be chosen. As for M and X , they have the same interpretation as earlier, namely as respectively the mass parameter and the dynamical variables of the limiting system. Setting $\omega_1 = -i\pi$, the contribution of each root α to the CM potential can be expressed as

$$m^2 \wp(\alpha \cdot x) = \frac{1}{2} M^2 \sum_{n=-\infty}^{\infty} \frac{e^{2\delta\omega_2}}{\text{ch}(\alpha \cdot x - 2n\omega_2) - 1}. \quad (4.279)$$

It suffices to consider positive roots α . We shall also assume that $0 \leq \delta \alpha \cdot \rho^\vee \leq 1$. The contributions of the $n = 0$ and $n = -1$ summands in

(4.279) are proportional to $e^{2\omega_2(\delta-\delta\alpha\cdot\rho^\vee)}$ and $e^{2\omega_2(\delta-1+\delta\alpha\cdot\rho^\vee)}$, respectively. Thus the existence of a finite scaling limit requires that

$$\delta \leq \delta\alpha \cdot \rho^\vee \leq 1 - \delta. \quad (4.280)$$

Let α_i , $1 \leq i \leq r$ be a basis of simple roots for \mathfrak{g} . If we want all simple roots α_i to survive in the limit, we must require that

$$\alpha_i \cdot \rho^\vee = 1, \quad 1 \leq i \leq r.$$

This condition characterizes the vector ρ^\vee as the *level vector*. Next, the second condition in (4.274) can be rewritten as $\delta\{1 + \max_\alpha(\alpha \cdot \rho^\vee)\} \leq 1$. But

$$h_{\mathfrak{g}} = 1 + \max_\alpha(\alpha \cdot \rho^\vee) \quad (4.281)$$

is precisely the Coxeter number of \mathfrak{g} , and we must have $\delta \leq \frac{1}{h_{\mathfrak{g}}}$. Thus when $\delta < \frac{1}{h_{\mathfrak{g}}}$, the contributions of all the roots except for the simple roots of \mathfrak{g} tend to 0. On the other hand, when $\delta = \frac{1}{h_{\mathfrak{g}}}$, the highest root α_0 realizing the maximum over α in (4.281) survives. Since $-\alpha_0$ is the additional simple root for the affine Lie algebra $\mathfrak{g}^{(1)}$, we arrive in this way at the following Theorem:

Under the limit (4.278), with $\delta = \frac{1}{h_{\mathfrak{g}}}$, and ρ^\vee given by the level vector, the Hamiltonian of the elliptic CM-system for the simple Lie algebra \mathfrak{g} tends to the Hamiltonian of the Toda system for the affine Lie algebra $\mathfrak{g}^{(1)}$.

Scaling Limit based on the Dual Coxeter Number

If the *SW-spectral curve* of the $N = 2$ supersymmetric gauge theory with a hypermultiplet in the adjoint representation is to be realized as the spectral curve for a CM-system, the parameter m in the CM-system should correspond to the mass of the hypermultiplet. In the gauge theory, the dependence of the coupling constant on the mass m is given by

$$\tau = \frac{i}{2\pi} h_{\mathfrak{g}}^\vee \ln \frac{m^2}{M^2} \quad \Longleftrightarrow \quad m = M q^{-\frac{1}{2h_{\mathfrak{g}}^\vee}}, \quad (4.282)$$

where $h_{\mathfrak{g}}^\vee$ is the quadratic *Casimir function* of the Lie algebra \mathfrak{g} . This shows that the correct physical limit, expressing the decoupling of the hypermultiplet as it becomes infinitely massive, is given by (4.271), but with $\delta = \frac{1}{h_{\mathfrak{g}}^\vee}$. To establish a closer parallel with our preceding discussion, we recall that

the quadratic Casimir $h_{\mathfrak{g}}^{\vee}$ coincides with the *dual Coxeter number* of \mathfrak{g} , defined by

$$h_{\mathfrak{g}}^{\vee} = 1 + \max_{\alpha} (\alpha^{\vee} \cdot \rho), \quad (4.283)$$

where $\alpha^{\vee} = \frac{2\alpha}{\alpha^2}$ is the coroot associated to α , and $\rho = \frac{1}{2} \sum_{\alpha > 0} \alpha$ is the well-known *Weyl vector*.

For simply laced Lie algebras \mathfrak{g} (ADE algebras), we have $h_{\mathfrak{g}} = h_{\mathfrak{g}}^{\vee}$, and the preceding scaling limits apply. However, for non simply-laced algebras (B_n, C_n, G_2, F_4), we have $h_{\mathfrak{g}} > h_{\mathfrak{g}}^{\vee}$, and our earlier considerations show that the untwisted elliptic CM Hamiltonians do not tend to a finite limit, $q \rightarrow 0$, M is kept fixed. This is why the twisted Hamiltonian systems (4.273) have to be introduced. The twisting produces precisely to an improvement in the asymptotic behavior of the potential which allows a finite, non-trivial limit. More precisely, we can write

$$m^2 \wp_{\nu}(x) = \frac{\nu^2}{2} \sum_{n=-\infty}^{\infty} \frac{m^2}{\operatorname{ch} \nu(x - 2n\omega_2) - 1}. \quad (4.284)$$

We have the following Theorem:

Under the limit

$$x = X + 2\omega_2 \frac{1}{h_{\mathfrak{g}}^{\vee}} \rho, \quad m = Mq^{-\frac{1}{2h_{\mathfrak{g}}^{\vee}}},$$

with ρ the Weyl vector and $q \rightarrow 0$, the Hamiltonian of the twisted elliptic CM-system for the simple Lie algebra \mathfrak{g} tends to the Hamiltonian of the Toda system for the affine Lie algebra $(\mathfrak{g}^{(1)})^{\vee}$.

So far we have discussed only the scaling limits of the Hamiltonians. However, similar arguments show that the Lax pairs constructed below also have finite, non-trivial scaling limits whenever this is the case for the Hamiltonians. The spectral parameter z should scale as $e^z = Zq^{\frac{1}{2}}$, with Z fixed. The parameter Z can be identified with the loop group parameter for the resulting affine Toda system.

4.14.11.5 Lax Pairs for CM-Systems

Let the rank of \mathfrak{g} be n , and d be its dimension. Let Λ be a representation of \mathfrak{g} of dimension N , of weights λ_I , $1 \leq I \leq N$. Let $u_I \in \mathbb{C}^N$ be the weights of the fundamental representation of $GL(N, \mathbb{C})$. Project orthogonally the u_I 's onto the λ_I 's as $su_I = \lambda_I + u_I$ and $\lambda_I \perp v_J$. It is easily verified that s^2 is the second *Dynkin index*. Then $\alpha_{IJ} = \lambda_I - \lambda_J$ is a weight of

$\Lambda \otimes \Lambda^*$ associated to the root $u_I - u_J$ of $GL(N, \mathbb{C})$. The Lax pairs for both untwisted and twisted CM-systems will be of the form

$$L = P + X, \quad M = D + X,$$

where the matrices P, X, D , and Y are given by

$$X = \sum_{I \neq J} C_{IJ} \Phi_{IJ}(\alpha_{IJ}, z) E_{IJ}, \quad Y = \sum_{I \neq j} C_{IJ} \Phi'_{IJ}(\alpha_{IJ}, z) E_{IJ}$$

and by $P = p \cdot h, \quad D = d \cdot (h \oplus \tilde{h}) + \Delta.$

Here h is in a Cartan subalgebra $\mathcal{H}_{\mathfrak{g}}$ for \mathfrak{g} , \tilde{h} is in the Cartan-Killing orthogonal complement of $\mathcal{H}_{\mathfrak{g}}$ inside a Cartan subalgebra \mathcal{H} for $GL(N, \mathbb{C})$, and Δ is in the centralizer of $\mathcal{H}_{\mathfrak{g}}$ in $GL(N, \mathbb{C})$. The functions $\Phi_{IJ}(x, z)$ and the coefficients C_{IJ} are yet to be determined. We begin by stating the necessary and sufficient conditions for the pair $L(z), M(z)$ of (4.275) to be a Lax pair for the (twisted or untwisted) CM-systems. For this, it is convenient to introduce the following notation

$$\Phi_{IJ} = \Phi_{IJ}(\alpha_{IJ} \cdot x),$$

$$\wp'_{IJ} = \Phi_{IJ}(\alpha_{IJ} \cdot x, z) \Phi'_{JI}(-\alpha_{IJ} \cdot x, z) - \Phi_{IJ}(-\alpha_{IJ} \cdot x, z) \Phi'_{JI}(\alpha_{IJ} \cdot x, z).$$

Then the Lax equation $\dot{L}(z) = [L(z), M(z)]$ implies the CM-system if and only if the following three identities are satisfied ($K \neq I \neq J$)

$$C_{IJ} C_{JI} \wp'_{IJ} \alpha_{IJ} = s^2 \sum_{\alpha \in \mathcal{R}(\mathfrak{g})} m_{|\alpha|}^2 \wp_{\nu(\alpha)}(\alpha \cdot x), \quad C_{IJ} C_{JI} \wp'_{IJ}(v_I - v_J) = 0,$$

$$C_{IK} C_{KJ} (\Phi_{IK} \Phi'_{KJ} - \Phi'_{IK} \Phi_{KJ}) =$$

$$s C_{IJ} \Phi_{IJ} d \cdot (v_I - v_J) + \Delta_{IJ} C_{KJ} \Phi_{KJ} - C_{IK} \Phi_{IK} \Delta_{KJ}$$

We have the following Theorem [D'Hoker and Phong (1998b)]:

A representation Λ , functions Φ_{IJ} , and coefficients C_{IJ} with a spectral parameter z satisfying (4.280–4.281) can be found for all twisted and untwisted elliptic CM-systems associated with a simple Lie algebra \mathfrak{g} , except possibly in the case of twisted G_2 . In the case of E_8 , we have to assume the existence of a ± 1 cocycle.

Lax Pairs for Untwisted CM Systems

Here are some important features of the Lax pairs obtained in this manner [D'Hoker and Phong (1998a); D'Hoker and Phong (1998b); D'Hoker and Phong (1998c)]:

In the case of the *untwisted* CM-systems, we can choose $\Phi_{IJ}(x, z) = \Phi(x, z)$, $\varphi_{IJ}(x) = \varphi(x)$ for all \mathfrak{g} .

$\Delta = 0$ for all \mathfrak{g} , except for E_8 .

For A_n , the Lax pair (4.269–4.270) corresponds to the choice of the fundamental representation for Λ . A different Lax pair can be found by taking Λ to be the antisymmetric representation.

For the BC_n system, the Lax pair is obtained by imbedding B_n in $GL(N, \mathbb{C})$ with $N = 2n + 1$. When $z = \omega_a$ (half-period), the Lax pair obtained this way reduces to the Lax pair obtained in [Olshanetsky and Perelomov (1976); Olshanetsky and Perelomov (1976)].

For the B_n and D_n systems, additional Lax pairs with spectral parameter can be found by taking Λ to be the spinor representation.

For G_2 , a first Lax pair with spectral parameter can be obtained by the above construction with Λ chosen to be the **7** of G_2 . A second Lax pair with spectral parameter can be obtained by restricting the **8** of B_3 to the $\mathbf{7} \oplus \mathbf{1}$ of G_2 .

For F_4 , a Lax pair can be obtained by taking Λ to be the $\mathbf{26} \oplus \mathbf{1}$ of F_4 , viewed as the restriction of the **27** of E_6 to its F_4 subalgebra.

For E_6 , Λ is the **27** representation.

Lax Pairs for Twisted CM Systems

Recall that the twisted and untwisted CM-systems differ only for non-simply laced Lie algebras, namely B_n , C_n , G_2 and F_4 . These are the only algebras we discuss in this paragraph. The construction (4.277–4.281) gives then Lax pairs for all of them, with the possible exception of twisted G_2 . Unlike the case of untwisted Lie algebras however, the functions Φ_{IJ} have to be chosen with care, and differ for each algebra.

For B_n , the Lax pair is of dimension $N = 2n$, admits two independent couplings m_1 and m_2 , and

$$\Phi_{IJ}(x, z) = \begin{cases} \Phi(x, z), & \text{if } I - J \neq 0, \pm n; \\ \Phi_2(\frac{1}{2}x, z), & \text{if } I - J = \pm n. \end{cases}$$

Here a new function $\Phi_2(x, z)$ is defined by

$$\Phi_2(\frac{1}{2}x, z) = \frac{\Phi(\frac{1}{2}x, z)\Phi(\frac{1}{2}x + \omega_1, z)}{\Phi(\omega_1, z)}.$$

For C_n , the Lax pair is of dimension $N = 2n + 2$, admits one independent

coupling m_2 , and

$$\Phi_{IJ}(x, z) = \Phi_2(x + \omega_{IJ}, z),$$

where ω_{IJ} are given by

$$\omega_{IJ} = \begin{cases} 0, & \text{if } I \neq J = 1, \dots, 2n+1; \\ \omega_2, & \text{if } 1 \leq I \leq 2n, J = 2n+2; \\ -\omega_2, & \text{if } 1 \leq J \leq 2n, I = 2n+2. \end{cases}$$

For F_4 , the Lax pair is of dimension $N = 24$, two independent couplings m_1 and m_2 ,

$$\Phi_{\lambda\mu}(x, z) = \begin{cases} \Phi(x, z), & \text{if } \lambda \cdot \mu = 0; \\ \Phi_1(\frac{1}{2}x, z), & \text{if } \lambda \cdot \mu = \frac{1}{2}; \\ \Phi_2(\frac{1}{2}x, z), & \text{if } \lambda \cdot \mu = -1. \end{cases}$$

where the function $\Phi_1(x, z)$ is defined by

$$\Phi_1(x, z) = \Phi(x, z) - e^{\pi i \zeta(z) + \eta_1 z} \Phi(x + \omega_1, z).$$

Here it is more convenient to label the entries of the Lax pair directly by the weights $\lambda = \lambda_I$ and $\mu = \lambda_J$ instead of I and J .

4.14.11.6 CM and SW Theory for $SU(N)$

The correspondence between SW theory for $N = 2$ super-YM theory with one hypermultiplet in the adjoint representation of the gauge algebra, and the elliptic CM-systems was first established in [D'Hoker and Phong (1998a)], for the gauge algebra $\mathfrak{g} = SU(N)$. We describe it here in some detail (see also [D'Hoker and Phong (1998b); D'Hoker and Phong (1998c)]).

All that we shall need here of the elliptic CM-system is its Lax operator $L(z)$, whose $N \times N$ matrix elements are given by

$$L_{ij}(z) = p_i \delta_{ij} - m(1 - \delta_{ij}) \Phi(x_i - x_j, z). \quad (4.285)$$

Notice that the Hamiltonian is simply given in terms of L by $H(x, p) = \frac{1}{2} \text{Tr} L(z)^2 + C \wp(z)$ with $C = -\frac{1}{2} m^2 N(N-1)$.

The correspondence between the data of the elliptic CM-system and those of the SW theory is as follows:

- (i) The parameter m in (4.285) is the hypermultiplet mass;

(ii) The gauge coupling g and the θ -angle are related to the modulus of the torus $\Sigma = \mathbb{C}/(2\omega_1\mathbb{Z} + 2\omega_2\mathbb{Z})$ by

$$\tau = \frac{\omega_2}{\omega_1} = \frac{\theta}{2\pi} + \frac{4\pi i}{g^2};$$

(iii) The SW curve Γ is the spectral curve of the elliptic CM model, defined by

$$\Gamma = \{(k, z) \in \mathbb{C} \times \Sigma, \det(kI - L(z)) = 0\}$$

and the SW 1-orm is $d\lambda = k \, dz$. Γ is invariant under the Weyl group of $SU(N)$.

(iv) Using the *Lax equation* $\dot{L} = [L, M]$, it is clear that the spectral curve is independent of time, and can be dependent only upon the constants of motion of the CM-system, of which there are only N . These integrals of motion may be viewed as parametrized by the quantum moduli of the SW system.

(v) Finally, $d\lambda = kdz$ is meromorphic, with a simple pole on each of the N sheets above the point $z = 0$ on the base torus. The residue at each of these poles is proportional to m , as required by the general set-up of SW theory.

Four Fundamental Theorems

1. The spectral curve equation $\det(kI - L(z)) = 0$ is equivalent to

$$\vartheta_1 \left(\frac{1}{2\omega_1} (z - m \frac{\partial}{\partial k}) | \tau \right) H(k) = 0,$$

where $H(k)$ is a monic polynomial in k of degree N , whose zeros (or equivalently whose coefficients) correspond to the moduli of the gauge theory. If $H(k) = \prod_{i=1}^N (k - k_i)$, then

$$\lim_{q \rightarrow 0} \frac{1}{2\pi i} \oint_{A_i} k dz = k_i - \frac{1}{2}m.$$

Here, ϑ_1 is the Jacobi ϑ -function, which admits a simple series expansion in powers of the instanton factor $q = e^{2\pi i \tau}$, so that the curve equation may also be rewritten as a series expansion

$$\sum_{n \in \mathbb{Z}} (-)^n q^{\frac{1}{2}n(n-1)} e^{nz} H(k - n \cdot m) = 0, \quad (4.286)$$

where we have set $\omega_1 = -i\pi$ without loss of generality. The series expansion (4.286) is superconvergent and sparse in the sense that it receives contributions only at integers that grow like n^2 .

2. The prepotential of the SW theory obeys a renormalization group-type equation that simply relates \mathcal{F} to the CM Hamiltonian, expressed in terms of the quantum order parameters a_j

$$a_j = \frac{1}{2\pi i} \oint_{A_j} d\lambda, \quad \frac{\partial \mathcal{F}}{\partial \tau} |_{a_j} = H(x, p) = \frac{1}{2} \text{Tr} L(z)^2 + C\wp(z). \quad (4.287)$$

Furthermore, in an expansion in powers of the instanton factor $q = e^{2\pi i \tau}$, the quantum order parameters a_j may be computed by residue methods in terms of the zeros of $H(k)$. The proof of (4.287) requires Riemann surface deformation theory. The fact that the quantum order parameters may be evaluated by residue methods arises from the fact that A_j -cycles may be chosen on the spectral curve Γ in such a way that they will shrink to zero as $q \rightarrow 0$. As a result, contour integrals around full-fledged branch cuts A_j reduce to contour integrals around poles at single points, which may be calculated by residue methods only. Knowing the quantum order parameters in terms of the zeros k_j of $H(k) = 0$ is a relation that may be inverted and used in (4.287) to get a differential relation for all order instanton corrections. It is now only necessary to evaluate explicitly the τ -independent contribution to \mathcal{F} , which in field theory arises from perturbation theory. This may be done easily by retaining only the $n = 0$ and $n = 1$ terms in the expansion of the curve (4.286), so that $z = \ln H(k) - \ln H(k - m)$. The results of the calculations to two instanton order may be summarized in the following Theorem:

3. The prepotential, to 2 instanton order is given by $\mathcal{F} = \mathcal{F}^{pert} + \mathcal{F}^{(1)} + \mathcal{F}^{(2)}$. The perturbative contribution is given by

$$\mathcal{F}^{pert} = \frac{\tau}{2} \sum_i a_i^2 - \frac{1}{8\pi i} \sum_{i,j} [(a_i - a_j)^2 \ln(a_i - a_j)^2 - (a_i - a_j - m)^2 \ln(a_i - a_j - m)^2] \quad (4.288)$$

while all instanton corrections are expressed in terms of a single function

$$S_i(a) = \frac{\prod_{j=1}^N [(a_i - a_j)^2 - m^2]}{\prod_{j \neq i} (a_i - a_j)^2}, \quad \text{as follows:}$$

$$\mathcal{F}^{(1)} = \frac{q}{2\pi i} \sum_i S_i(a_i), \quad (4.289)$$

$$\mathcal{F}^{(2)} = \frac{q^2}{8\pi i} \left[\sum_i S_i(a_i) \partial_i^2 S_i(a_i) + 4 \sum_{i \neq j} \frac{S_i(a_i) S_j(a_j)}{(a_i - a_j)^2} - \frac{S_i(a_i) S_j(a_j)}{(a_i - a_j - m)^2} \right].$$

The perturbative corrections to the prepotential of (4.288) indeed precisely agree with the predictions of asymptotic freedom. The formulas (4.289) for the instanton corrections $\mathcal{F}^{(1)}$ and $\mathcal{F}^{(2)}$ are new, as they have not yet been computed by direct field theory methods. The moduli k_i , $1 \leq i \leq N$, of the gauge theory are evidently integrals of motion of the system. To identify these integrals of motion, denote by S be any subset of $\{1, \dots, N\}$, and let $S^* = \{1, \dots, N\} \setminus S$, $\wp(S) = \wp(x_i - x_j)$ when $S = \{i, j\}$. Let also p_S denote the subset of momenta p_i with $i \in S$.

4. For any K , $0 \leq K \leq N$, let $\sigma_K(k_1, \dots, k_N) = \sigma_K(k)$ be the K -th symmetric polynomial of (k_1, \dots, k_N) , defined by

$$H(u) = \sum_{K=0}^N (-)^K \sigma_K(k) u^{N-K}. \quad \text{Then}$$

$$\sigma_K(k) = \sigma_K(p) + \sum_{l=1}^{[K/2]} m^{2l} \sum_{\substack{|S_i \cap S_j| = 2\delta_{ij} \\ 1 \leq i, j \leq l}} \sigma_{K-2l}(p_{(\cup_{i=1}^l S_i)^*}) \prod_{i=1}^l [\wp(S_i) + \frac{\eta_1}{\omega_1}].$$

4.14.11.7 CM and SW Theory for General Lie Algebra

Now, we consider the $N = 2$ supersymmetric gauge theory for a general simple gauge algebra \mathfrak{g} and a hypermultiplet of mass m in the adjoint representation. Then we have the following results [D'Hoker and Phong (1998d)].

The SW curve of the theory is given by the spectral curve

$$\Gamma = \{(k, z) \in \mathbb{C} \times \Sigma; \det(kI - L(z)) = 0\}$$

of the *twisted* elliptic CM-system associated to the Lie algebra \mathfrak{g} . The SW differential $d\lambda$ is given by $d\lambda = kdz$.

The function $R(k, z) = \det(kI - L(z))$ is polynomial in k and meromorphic in z . The spectral curve Γ is invariant under the Weyl group of \mathfrak{g} . It depends on n complex moduli, which can be thought of as independent integrals of motion of the CM-system.

The differential $d\lambda = kdz$ is meromorphic on Γ , with simple poles.

The position and residues of the poles are independent of the moduli. The residues are linear in the hypermultiplet mass m (unlike the case of $SU(N)$, their exact values are difficult to determine for general \mathfrak{g} .)

In the $m \rightarrow 0$ limit, the CM-system reduces to a free system, the spectral curve Γ is just the product of several unglued copies of the base torus Σ , indexed by the constant eigenvalues of $L(z) = p \cdot h$. Let k_i , $1 \leq i \leq n$, be n independent eigenvalues, and A_i, B_i be the A and B cycles lifted to the corresponding sheets. For each i , we readily get

$$\begin{aligned} a_i &= \frac{1}{2\pi i} \oint_{A_i} d\lambda = \frac{k_i}{2\pi i} \oint_A dz = \frac{2\omega_1}{2\pi i} k_i, \\ a_{Di} &= \frac{1}{2\pi i} \oint_{B_i} d\lambda = \frac{k_i}{2\pi i} \oint_B dz = \frac{2\omega_1}{2\pi i} \tau k_i. \end{aligned}$$

Thus the prepotential \mathcal{F} is given by $\mathcal{F} = \frac{\tau}{2} \sum_{i=1}^n a_i^2$. This is the classical prepotential and hence the correct answer, since in the $m \rightarrow 0$ limit, the theory acquires an $N = 4$ supersymmetry, and receives no quantum corrections.

The $m \rightarrow \infty$ limit is the crucial consistency check, which motivated the introduction of the *twisted* CM-systems in the first place. In the limit $m \rightarrow \infty$, $q \rightarrow 0$, with

$$x = X + 2\omega_2 \frac{1}{h_{\mathfrak{g}}^{\vee}} \rho, \quad m = Mq^{-\frac{1}{2h_{\mathfrak{g}}^{\vee}}},$$

with X and M kept fixed, the Hamiltonian and spectral curve for the twisted elliptic CM-system with Lie algebra \mathfrak{g} reduce to the Hamiltonian and spectral curve for the Toda system for the affine Lie algebra $(\mathfrak{g}^{(1)})^{\vee}$. This is the correct answer. Indeed, in this limit, the gauge theory with adjoint hypermultiplet reduces to the pure YM theory, and the SW spectral curves for pure YM with gauge algebra \mathfrak{g} have been shown to be the spectral curves of the Toda system for $(\mathfrak{g}^{(1)})^{\vee}$. The effective prepotential can be evaluated explicitly in the case of $\mathfrak{g} = D_n$ for $n \leq 5$. Its logarithmic singularity does reproduce the logarithmic singularities expected from field theory considerations.

As in the known correspondences between SW theory and integrable models, we expect the following equation to hold [D'Hoker and Phong (1998d)]

$$\frac{\partial \mathcal{F}}{\partial \tau} = H_{\mathfrak{g}}^{\text{twisted}}(x, p).$$

Note that the left hand side can be interpreted in the gauge theory as a renormalization group equation.

For simple laced \mathfrak{g} , the curves $R(k, z) = 0$ are modular invariant. Physically, the gauge theories for these Lie algebras are self-dual. For non simply-laced \mathfrak{g} , the modular group is broken to the congruence subgroup $\Gamma_0(2)$ for $\mathfrak{g} = B_n, C_n, F_4$, and to $\Gamma_0(3)$ for G_2 . The Hamiltonians of the twisted CM-systems for non-simply laced \mathfrak{g} are also transformed under Landen transformations into the Hamiltonians of the twisted CM-system for the dual algebra \mathfrak{g}^\vee . It would be interesting to determine whether such transformations exist for the spectral curves or the corresponding gauge theories themselves.

4.14.12 SW Theory and WDVV Equations

As presented above, N. Seiberg and E. Witten proposed in [Seiberg and Witten (1994a); Seiberg and Witten (1994b)] a new way to deal with the low-energy effective actions of $N = 2$ 4D supersymmetric gauge theories, both pure gauge theories (i.e., containing only vector super-multiplet) and those with matter hypermultiplets. Among other things, they have shown that the low-energy effective actions (the end-points of the renormalization group flows) fit into universality classes depending on the vacuum of the theory. If the moduli space of these vacua is a finite-dimensional variety, the effective actions can be essentially described in terms of system with *finite*-dimensional phase space (number of degrees of freedom is equal to the rank of the gauge group), although the original theory lives in a many-dimensional space-time.

4.14.12.1 WDVV Equations

Now, it turns out that the prepotential of SW effective theory satisfies the following set of *Witten-Dijkgraaf-Verlinde-Verlinde* (WDVV) equations:

$$\mathcal{F}_i \mathcal{F}_k^{-1} \mathcal{F}_j = \mathcal{F}_j \mathcal{F}_k^{-1} \mathcal{F}_i \quad (i, j, k = 1, \dots, n), \quad (4.290)$$

where \mathcal{F}_i are the matrices on a moduli space \mathcal{M} of third order derivatives

$$(\mathcal{F}_i)_{jk} = \frac{\partial^3 \mathcal{F}}{\partial a_i \partial a_j \partial a_k}$$

of a function $\mathcal{F}(a_1, \dots, a_n)$, with the prepotential variables a_i .

Although generally there is a lot of solutions to the matrix equations (4.290), it is extremely non-trivial task to express all the matrix elements through the only function \mathcal{F} . In fact, there have been only known the two different classes of the non-trivial solutions to the WDVV equations, both being intimately related to the 2D topological theories of type A (quantum cohomologies) and of type B ($N = 2$ SUSY Landau-Ginzburg (LG) theories). Thus, the existence of a new class of solutions connected with the 4D theories looks quite striking. It is worth noting that both the 2D topological theories and the SW theories reveal the integrability structures related to the WDVV equations.

This system of nonlinear equations is satisfied by the SW prepotential defining the low-energy effective action. Moreover the leading perturbative approximation to this exact SW prepotential should satisfy this set of equations by itself. For instance, for the gauge group $SU(n)$ the expression [Mironov (1998)]

$$\mathcal{F}^{pert} = \frac{1}{4} \sum_{i \leq j \leq n-1} (a_i - a_j)^2 \log(a_i - a_j)^2 + \frac{1}{2} \sum_{i=1}^{n-1} a_i^2 \log a_i^2$$

defines a solution of the generalized WDVV-system (4.290).

Clearly, other gauge groups may be considered and more general solutions may be proposed for classical Lie groups. So although extremely difficult to solve in general, this overdetermined system of nonlinear equations admit exact solutions. Martini [Martini and Gragert (1999)] proved that a substantial class of solutions for the system (4.290) could be constructed from root systems of semisimple Lie algebras. Precisely, let R be the root system of a semisimple Lie algebra \mathfrak{g} . Then the function

$$\mathcal{F} = \frac{1}{4} \sum_{\alpha \in R} (\alpha, a)^2 \log(\alpha, a)^2,$$

defined on the Cartan subalgebra \mathfrak{h} of \mathfrak{g} satisfies the generalized WDVV equations (4.290). Here the bracket represents the *Killing form* of \mathfrak{g} .

Now, to give some more insight of the general structure of the WDVV equations, let us consider the simplest non-trivial examples of $n = 3$ WDVV equations in topological theories. The first example is the $N = 2$ SUSY LG theory with the superpotential $W'(\lambda) = \lambda^3 - q$. In this case, the prepotential reads as

$$\mathcal{F} = \frac{1}{2} a_1 a_2^2 + \frac{1}{2} a_1^2 a_3 + \frac{q}{2} a_2 a_3^2 \quad (4.291)$$

and the matrices \mathcal{F}_i (the third derivatives of the prepotential) are

$$\mathcal{F}_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \mathcal{F}_2 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & q \end{pmatrix}, \quad \mathcal{F}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & q \\ 0 & q & 0 \end{pmatrix}.$$

The second example is the quantum cohomologies of CP^2 . In this case, the prepotential is given by the formula

$$\mathcal{F} = \frac{1}{2}a_1a_2^2 + \frac{1}{2}a_1^2a_3 + \sum_{k=1}^{\infty} \frac{N_k a_3^{3k-1}}{(3k-1)!} e^{ka_2}, \quad (4.292)$$

where the coefficients N_k (describing the rational Gromov–Witten classes) count the number of the rational curves in CP^2 and are to be calculated. Since the matrices \mathcal{F}_i have the form

$$\mathcal{F}_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \mathcal{F}_2 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & \mathcal{F}_{222} & \mathcal{F}_{223} \\ 0 & \mathcal{F}_{223} & \mathcal{F}_{233} \end{pmatrix}, \quad \mathcal{F}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \mathcal{F}_{223} & \mathcal{F}_{233} \\ 0 & \mathcal{F}_{233} & \mathcal{F}_{333} \end{pmatrix},$$

the WDVV equations are equivalent to the identity [Mironov (1998)]

$$\mathcal{F}_{333} = \mathcal{F}_{223}^2 - \mathcal{F}_{222}\mathcal{F}_{233},$$

which, in turn, results into the recurrent relation defining the coefficients N_k :

$$\frac{N_k}{(3k-4)!} = \sum_{a+b=k} \frac{a^2b(3b-1)b(2a-b)}{(3a-1)!(3b-1)!} N_a N_b.$$

The crucial feature of the presented examples is that, in both cases, there exists a constant matrix \mathcal{F}_1 . One can consider it as a flat metric on the moduli space. In fact, in its original version, the WDVV equations have been written in a slightly different form, that is, as the associativity condition of some algebra. Having distinguished the (constant) metric $\eta \equiv \mathcal{F}_1$, one can naturally rewrite (4.290) as the equations

$$C_i C_j = C_j C_i \quad (4.293)$$

for the matrices $(C_i)_{jk} \equiv \eta^{-1}\mathcal{F}_i$, i.e., $C_{ik}^j = \eta^{jl}\mathcal{F}_{ilk}$. Formula (4.293) is equivalent to (4.290) with $j = 1$. Moreover, this particular relation is already sufficient to reproduce the whole set of the WDVV equations (4.290).

Let us also note that, although the WDVV equations can be fulfilled only for some specific choices of the coordinates a_i on the moduli space, they still admit any linear transformation. This defines the flat structures on the moduli space, and we often call a_i flat coordinates.

4.14.12.2 Perturbative SW Prepotentials

Before going into the discussion of the WDVV equations for the complete SW prepotentials, let us note that the leading perturbative part of them should satisfy the equations (4.290) by itself (since the classical quadratic piece does not contribute into the third derivatives). On the other hand, if the WDVV equations are fulfilled for perturbative prepotential, it is a necessary condition for them to hold for complete prepotential.

To determine the one-loop perturbative prepotential from the field theory calculation, let us note that there are two origins of masses in $\mathcal{N} = 2$ SUSY YM models: first, they can be generated by vacuum values of the scalar ϕ from the gauge super-multiplet. For a super-multiplet in representation R of the gauge group G this contribution to the prepotential is given by the analog of the Coleman-Weinberg formula (from now on, we omit the classical part of the prepotential from all expressions):

$$\mathcal{F}_R = \pm \frac{1}{4} \text{Tr}_R \phi^2 \log \phi, \quad (4.294)$$

and the sign is ‘+’ for vector super-multiplets (normally they are in the adjoint representation) and ‘−’ for matter hypermultiplets. Second, there are bare masses m_R which should be added to ϕ in (4.294). As a result, the general expression for the perturbative prepotential is

$$\begin{aligned} \mathcal{F} = & \frac{1}{4} \sum_{\substack{\text{vector} \\ \text{multiplets}}} \text{Tr}_A(\phi + M_n I_A)^2 \log(\phi + M_n I_A) - \\ & - \frac{1}{4} \sum_{\substack{\text{hyper} \\ \text{multiplets}}} \text{Tr}_R(\phi + m_R I_R)^2 \log(\phi + m_R I_R) + f(m), \end{aligned} \quad (4.295)$$

where the term $f(m)$, depending only on masses, is not fixed by the (perturbative) field theory, but can be read off from the non-perturbative description, and I_R denotes the unit matrix in the representation R .

As an example, consider the $SU(n)$ gauge group. Then, e.g., perturba-

tive prepotential for the pure gauge theory acquires the form

$$\mathcal{F}_V^{pert} = \frac{1}{4} \sum_{ij} (a_i - a_j)^2 \log(a_i - a_j).$$

This formula states that when eigenvalues of the scalar fields in the gauge super-multiplet are non-vanishing (perturbatively a_r are eigenvalues of the vacuum expectation matrix $\langle \phi \rangle$), the fields in the gauge multiplet acquire masses $m_{rr'} = a_r - a_{r'}$ (the pair of indices (r, r') label a field in the adjoint representation of G). In the $SU(n)$ case, the eigenvalues are subject to the condition $\sum_i a_i = 0$. Analogous formulas for the adjoint matter contribution to the prepotential is

$$\mathcal{F}_A^{pert} = -\frac{1}{4} \sum_{ij} (a_i - a_j + M)^2 \log(a_i - a_j + M),$$

while the contribution of the fundamental matter reads

$$\mathcal{F}_{\mathcal{F}}^{pert} = -\frac{1}{4} \sum_i (a_i + m)^2 \log(a_i + m).$$

The perturbative prepotentials have the following characteristics [Mironov (1998)]:

(i) The WDVV equations always hold for the pure gauge theories:
 $\mathcal{F}^{pert} = \mathcal{F}_V^{pert}$.

(ii) If one considers the gauge super-multiplets interacting with the matter hypermultiplets in the first fundamental representation with masses m_α $\mathcal{F}^{pert} = \mathcal{F}_V^{pert} + r\mathcal{F}_{\mathcal{F}}^{pert} + Kf_{\mathcal{F}}(m)$ (where r and K are some undetermined coefficients), the WDVV equations do not hold unless

$$K = r^2/4, \quad f_{\mathcal{F}}(m) = \frac{1}{4} \sum_{\alpha, \beta} (m_\alpha - m_\beta)^2 \log(m_\alpha - m_\beta),$$

the masses being regarded as moduli (i.e., the equations (4.290) contain the derivatives with respect to masses).

(iii) If in the theory the adjoint matter hypermultiplets are presented, i.e., $\mathcal{F}^{pert} = \mathcal{F}_V^{pert} + \mathcal{F}_A^{pert} + f_A(m)$, the WDVV equations never hold.

From the investigation of the WDVV equations for the perturbative prepotentials, one can learn the following lessons:

- Masses are to be regarded as moduli.
- As an empiric rule, one may say that the WDVV equations are satisfied by perturbative prepotentials which depend only on the pairwise sums

of the type $(a_i \pm b_j)$, where moduli a_i and b_j are either periods or masses³⁸. This is the case for the models that contain either massive matter hypermultiplets in the first fundamental representation (or its dual), or massless matter in the square product of those. Troubles arise in all other situations because of the terms with $a_i \pm b_j \pm c_k \pm \dots$

- At value $r = 2$, like a_i 's lying in irrep of G , masses m_α 's can be regarded as lying in irrep of some \tilde{G} so that if $G = A_n, C_n, D_n$, $\tilde{G} = A_n, D_n, C_n$ accordingly. This correspondence 'explains' the form of the mass term in the prepotential $f(m)$.

4.14.12.3 Associativity Conditions

In the context of the 2D LG topological theories, the WDVV equations arose as associativity condition of some polynomial algebra. Mironov has proved in [Mironov (1998)] that the equations in the SW theories have the same origin.

In this case, one deals with the chiral ring formed by a set of polynomials $\{\Phi_i(\lambda)\}$ and two co-prime (i.e., without common zeroes) fixed polynomials $Q(\lambda)$ and $P(\lambda)$. The polynomials Φ form the associative algebra with the structure constants C_{ij}^k given with respect to the product defined by modulo P' :

$$\Phi_i \Phi_j = C_{ij}^k \Phi_k Q' + (*)P' \longrightarrow C_{ij}^k \Phi_k Q', \quad (4.296)$$

the associativity condition being

$$(\Phi_i \Phi_j) \Phi_k = \Phi_i (\Phi_j \Phi_k), \quad (4.297)$$

$$\text{i.e., } C_i C_j = C_j C_i, \quad (C_i)_k^j = C_{ik}^j. \quad (4.298)$$

Now, in order to get from these conditions the WDVV equations, one needs to choose properly the flat moduli:

$$a_i = -\frac{n}{i(n-i)} \text{Res} \left(P^{i/n} dQ \right), \quad n = \text{ord}(P).$$

Then, there exists the prepotential whose third derivatives are given by the residue formula

$$\mathcal{F}_{ijk} = \frac{1}{2\pi i} \text{Res}_{P'=0} \frac{\Phi_i \Phi_j \Phi_k}{P'}. \quad (4.299)$$

³⁸This general rule can be easily interpreted in D-brane terms, since the interaction of branes is caused by strings between them. The pairwise structure $(a_i \pm b_j)$ exactly reflects this fact, a_i and b_j should be identified with the ends of string.

On the other hand, from the associativity condition (4.298) and residue formula (4.299), one obtains that

$$\mathcal{F}_{ijk} = (C_i)_j^l F_{Q'lk}, \quad \text{i.e.,} \quad C_i = \mathcal{F}_i F_{Q'}^{-1}. \quad (4.300)$$

Substituting this formula for C_i into (4.298), one finally reaches the equations of the WDVV type. The choice $Q' = \Phi_l$ gives the standard equations (4.290). In 2D topological theories, there is always the unity operator that corresponds to $Q' = 1$ and leads to the constant metric $\mathcal{F}_{Q'}$.

Thus, from this short study of the WDVV equations in the LG theories, we can get three main ingredients necessary for these equations to hold. These are:

- associative algebra
- flat moduli (coordinates)
- residue formula

In the SW theory only the first ingredient requires a non-trivial check.

4.14.12.4 SW Theories and Integrable Systems

Now we turn to the WDVV equations in the SW construction [Seiberg and Witten (1994a); Seiberg and Witten (1994b)] and show how they are related to an *integrable system* underlying the corresponding SW theory. The most important result of [Seiberg and Witten (1994a); Seiberg and Witten (1994b)], from this point of view, is that the moduli space of vacua and low energy effective action in SYM theories are completely given by the following input data:

- Riemann surface \mathcal{C}
- moduli space \mathcal{M} (of the curves \mathcal{C})
- meromorphic 1-form dS on \mathcal{C}

This input can be naturally described in the framework of some underlying integrable system. Let us consider a concrete example: the $SU(n)$ pure gauge SYM theory that can be described by the periodic *Toda chain* with n sites. This integrable system is entirely given by the *Lax operator*

$$L(w) = \begin{pmatrix} p_1 & e^{q_1 - q_2} & w \\ e^{q_1 - q_2} & p_2 & \vdots \\ & \dots & \ddots & \vdots \\ \frac{1}{w} & & \dots & p_n \end{pmatrix}. \quad (4.301)$$

The Riemann surface \mathcal{C} of the SW data is nothing but the spectral curve of the integrable system, which is given by the equation: $\det(L(w) - \lambda) = 0$. Taking into account (4.301), one can get from this formula the equation

$$w + \frac{1}{w} = P(\lambda) = \prod_{i=1}^n (\lambda - \lambda_i), \quad \sum_i \lambda_i = 0, \quad (4.302)$$

where the ramification points λ_i are Hamiltonians (i.e., integrals of motion) parametrizing the moduli space \mathcal{M} of the spectral curves. The substitution $Y \equiv w - 1/w$ transforms the curve (4.302) to the standard hyper-elliptic form $Y^2 = P^2 - 4$, the genus of the curve being $n - 1$.

As to the meromorphic 1-form

$$dS = \lambda \frac{dw}{w} = \lambda \frac{dP}{Y},$$

it is just the shorten action ‘ pdq ’ along the non-contractible contours on the Hamiltonian tori. Its defining property is that the derivatives of dS with respect to the moduli (ramification points) are holomorphic differentials on the spectral curve.

Now, following [Mironov (1998)], let us describe the general integrable framework for the SW construction and start with the theories without matter hypermultiplets. First, matter hypermultiplets. First, we introduce bare spectral curve E that is torus: $y^2 = x^3 + g_2 x^2 + g_3$ for the UV finite SYM theories with the associated *holomorphic 1-form*: $d\omega = dx/y$. This bare spectral curve degenerates into the double-punctured sphere (annulus) for the asymptotically free theories:

$x \rightarrow w + 1/w$, $y \rightarrow w - 1/w$, $d\omega = dw/w$. On this bare curve, there is given a matrix-valued Lax operator $L(x, y)$. The dressed spectral curve is defined from the formula $\det(L - \lambda) = 0$. This spectral curve is a ramified covering of E given by the equation

$$\mathcal{P}(\lambda; x, y) = 0 \quad (4.303)$$

In the case of the gauge group $G = SU(n)$, the function \mathcal{P} is a polynomial of degree n in λ .

Thus, the moduli space \mathcal{M} of the spectral curve is given just by coefficients of \mathcal{P} . The generating 1-form $dS \cong \lambda dw$ is meromorphic on \mathcal{C} (hereafter the equality modulo total derivatives is denoted by ‘ \cong ’).

The prepotential and other ‘physical’ quantities are defined in terms of the *cohomology class* of dS :

$$a_i = \oint_{A_i} dS, \quad a_i^D \equiv \frac{\partial \mathcal{F}}{\partial a_i} = \oint_{B_i} dS, \quad A_I \circ B_J = \delta_{IJ}. \tag{4.304}$$

The first identity defines here the appropriate flat moduli, while the second one – the prepotential. The derivatives of the generating differential dS give holomorphic 1-differentials:

$$\frac{\partial dS}{\partial a_i} = d\omega_i \tag{4.305}$$

and, therefore, the second derivative of the prepotential is the period matrix of the curve \mathcal{C} :

$$\frac{\partial^2 \mathcal{F}}{\partial a_i \partial a_j} = T_{ij}.$$

The latter formula allows one to identify prepotential with logarithm of the τ –function of Whitham hierarchy: $\mathcal{F} = \log \tau$.

So far we reckoned without massive hypermultiplets. In order to include them, one just needs to consider the surface \mathcal{C} with punctures. Then, the masses are proportional to residues of dS at the punctures, and the moduli space has to be extended to include these mass moduli. The correspondence between SYM theories and integrable systems is built through the SW construction in most of known cases that are collected in the following table [Mironov (1998)].

SUSY physical theory	4D pure gauge SYM theory, gauge group G	4D SYM with fundamental matter	4D SYM with adjoint matter	5d pure gauge SYM theory
Underlying integrable system	Toda chain for the dual affine \hat{G}^\vee	Rational spin chain of XXX type	Calogero–Moser system	Relativistic Toda chain
Bare spectral curve	sphere	sphere	torus	sphere
Dressed spectral curve	hyper-elliptic	hyper-elliptic	non-hyper-elliptic	hyper-elliptic
Generating meromorphic 1-form dS	$\lambda \frac{dw}{w}$	$\lambda \frac{dw}{w}$	$\lambda \frac{dx}{y}$	$\log \lambda \frac{dw}{w}$

Correspondence: SUSY gauge theories \iff integrable systems

To complete this table, we describe the dressed spectral curves in each case in more explicit terms. Let us note that in all but adjoint matter cases the curves are hyper-elliptic and can be described by the general formula

$$\mathcal{P}(\lambda, w) = 2P(\lambda) - w - \frac{\mathcal{Q}(\lambda)}{w}. \quad (4.306)$$

Here $P(\lambda)$ is characteristic polynomial of the algebra G itself, i.e.,

$$P(\lambda) = \det(G - \lambda I) = \prod_i (\lambda - \lambda_i),$$

where determinant is taken in the first fundamental representation and λ_i 's are the eigenvalues of the algebraic element G . For the pure gauge theories with the classical groups, $\mathcal{Q}(\lambda) = \lambda^{2s}$ and

$$A_{n-1} : P(\lambda) = \prod_{i=1}^n (\lambda - \lambda_i), \quad s = 0; \quad (4.307)$$

$$B_n : P(\lambda) = \lambda \prod_{i=1}^n (\lambda^2 - \lambda_i^2), \quad s = 2; \quad (4.308)$$

$$C_n : P(\lambda) = \prod_{i=1}^n (\lambda^2 - \lambda_i^2), \quad s = -2; \quad (4.309)$$

$$D_n : P(\lambda) = \prod_{i=1}^n (\lambda^2 - \lambda_i^2), \quad s = 2 \quad (4.310)$$

For exceptional groups, the curves arising from the characteristic polynomials of the dual affine algebras do not acquire the hyper-elliptic form. Therefore, in this case, the line 'dressed spectral curve' in the table has to be corrected.

In order to include $n_{\mathcal{F}}$ massive hypermultiplets in the first fundamental representation, one can just change λ^{2s} for $\mathcal{Q}(\lambda) = \lambda^{2s} \prod_{i=1}^{n_{\mathcal{F}}} (\lambda - m_i)$ if $G = A_n$ and for $\mathcal{Q}(\lambda) = \lambda^{2s} \prod_{i=1}^{n_{\mathcal{F}}} (\lambda^2 - m_i^2)$ if $G = B_n, C_n, D_n$.

At last, the 5D theory is just described by $\mathcal{Q}(\lambda) = \lambda^{n/2}$.

In the Calogero–Moser case, the spectral curve is non-hyper-elliptic, since the bare curve is elliptic. Therefore, it can be described as some covering of the hyper-elliptic curve.

4.14.12.5 WDVV Equations in SW Theories

As we already discussed, in order to derive the WDVV equations along the line used in the context of the LG theories, we need three crucial ingredients:

flat moduli, residue formula and associative algebra. However, the first two of these are always contained in the SW construction provided the underlying integrable system is known. Indeed, one can derive (see [4,5]) the following residue formula

$$\mathcal{F}_{ijk} = \operatorname{Res}_{d\omega=0} \frac{d\omega_i d\omega_j d\omega_k}{d\omega d\lambda}, \quad (4.311)$$

where the proper flat moduli a_i 's are given by formula (4.304). Thus, the only point is to be checked is the existence of the associative algebra. The residue formula (4.311) hints that this algebra is to be the algebra Ω^1 of the holomorphic differentials $d\omega_i$. In the forthcoming discussion we restrict ourselves to the case of pure gauge theory, the general case being treated in complete analogy.

Let us consider the algebra Ω^1 and fix three differentials $dQ, d\omega, d\lambda \in \Omega^1$. The product in this algebra is given by the expansion

$$d\omega_i d\omega_j = C_{ij}^k d\omega_k dQ + (*)d\omega + (*)d\lambda \quad (4.312)$$

that should be factorized over the ideal spanned by the differentials $d\omega$ and $d\lambda$. This product belongs to the space of quadratic holomorphic differentials:

$$\Omega^1 \cdot \Omega^1 \in \Omega^2 \cong \Omega^1 \cdot (dQ \oplus d\omega \oplus d\lambda). \quad (4.313)$$

Since the dimension of the space of quadratic holomorphic differentials is equal to $3g - 3$, the l.h.s. of (6.6) with arbitrary $d\omega_i$'s is the vector space of dimension $3g - 3$. At the same time, at the r.h.s. of (6.6) there are g arbitrary coefficients C_{ij}^k in the first term (since there are exactly so many holomorphic 1-forms that span the arbitrary holomorphic 1-form $C_{ij}^k d\omega_k$), $g - 1$ arbitrary holomorphic differentials in the second term (one differential should be subtracted to avoid the double counting) and $g - 2$ holomorphic 1-forms in the third one. Thus, totally we get that the r.h.s. of (6.6) is spanned also by the basis of dimension $g + (g - 1) + (g - 2) = 3g - 3$.

This means that the algebra exists in the general case of the SW construction. However, this algebra generally is not associative. This is because, unlike the LG case, when it was the algebra of polynomials and, therefore, the product of the two belonged to the same space (of polynomials), product in the algebra of holomorphic 1-differentials no longer belongs to the same space but to the space of quadratic holomorphic differentials. Indeed, to check associativity, one needs to consider the triple product of

Ω^1 :

$$\Omega^1 \cdot \Omega^1 \cdot \Omega^1 \in \Omega^3 = \Omega^1 \cdot (dQ)^2 \oplus \Omega^2 \cdot d\omega \oplus \Omega^2 \cdot d\lambda \quad (4.314)$$

Now let us repeat our calculation: the dimension of the l.h.s. of this expression is $5g - 5$ that is the dimension of the space of holomorphic 3-differentials. The dimension of the first space in expansion of the r.h.s. is g , the second one is $3g - 4$ and the third one is $2g - 4$. Since $g + (3g - 4) + (2g - 4) = 6g - 8$ is greater than $5g - 5$ (unless $g \leq 3$), formula (4.314) does not define the unique expansion of the triple product of Ω^1 and, therefore, the associativity spoils.

The situation can be improved if one considers the curves with additional involutions. As an example, let us consider the family of hyper-elliptic curves: $y^2 = \text{Pol}_{2g+2}(\lambda)$. In this case, there is the involution, $\sigma : y \rightarrow -y$ and Ω^1 is spanned by the σ -odd holomorphic 1-forms $\frac{x^{i-1}dx}{y}$, $i = 1, \dots, g$. Let us also note that both dQ and $d\omega$ are σ -odd, while $d\lambda$ is σ -even. This latter fact means that $d\lambda$ can be only meromorphic unless there are punctures on the surface (which is, indeed, the case in the presence of the mass hypermultiplets). Thus, formula (6.6) can be replaced by that without $d\lambda$

$$\Omega_+^2 = \Omega_-^1 \cdot dQ \oplus \Omega_-^1 \cdot d\omega, \quad (4.315)$$

where we have expanded the space of holomorphic 2-forms into the parts with definite σ -parity: $\Omega^2 = \Omega_+^2 \oplus \Omega_-^2$, which are manifestly given by the differentials $\frac{x^{i-1}(dx)^2}{y^2}$, $i = 1, \dots, 2g - 1$ and $\frac{x^{i-1}(dx)^2}{y}$, $i = 1, \dots, g - 2$ respectively. Now it is easy to understand that the dimensions of the l.h.s. and r.h.s. of (4.315) coincide and are equal to $2g - 1$.

Analogously, in this case, one can check the associativity. It is given by the expansion

$$\Omega_-^3 = \Omega_-^1 \cdot (dQ)^2 \oplus \Omega_+^2 \cdot d\omega,$$

where both the l.h.s. and r.h.s. have the same dimension: $3g - 2 = g + (2g - 2)$. Thus, the algebra of holomorphic 1-forms on hyper-elliptic curve is really associative [Mironov (1998)].

This page intentionally left blank

Chapter 5

Applied Jet Geometry

Modern formulation of generalized Lagrangian and Hamiltonian dynamics on fibre bundles is developed in the language of *jet spaces*, or *jet manifolds* (see [Kolar *et al.* (1993); Saunders (1989); Griffiths (1983); Bryant *et al.* (1991); Bryant *et al.* (2003); Giachetta *et al.* (1997); Mangiarotti *et al.* (1999); Mangiarotti and Sardanashvily (2000a); Saunders (1989); Sardanashvily (1993); Sardanashvily (1995); Sardanashvily (2002a)]).

Roughly speaking, given two smooth manifolds M and N , the two smooth maps $f, g : M \rightarrow N$ between them are said to determine the same k -jet at a point $x \in M$, if they have the k th order contact (or, the k th order tangency) at x [Kolar *et al.* (1993); Arnold (1988a)]. A set of all k -jets from M to N is a jet space $J^k(M, N)$. It is a generalization of a tangent bundle that makes a new smooth fiber bundle out of a given smooth fiber bundle – following the recursive n -categorical process. It makes it possible to write differential equations on sections of a fiber bundle in an invariant form. Historically, jet spaces are attributed to C. Ehresmann, and were an advance on the method of prolongation of E. Cartan, of dealing geometrically with higher derivatives, by imposing differential form conditions on newly-introduced formal variables.

5.1 Intuition Behind a Jet Space

The concept of *jet space* is based on the idea of *higher-order tangency*, or *higher-order contact*, at some designated point on a smooth manifold (see [Arnold (1988a); Kolar *et al.* (1993)]). Namely, a pair of smooth manifold

maps (see Figure 5.1),

$$f_1, f_2 : M \rightarrow N$$

are said to be k -tangent (or tangent of order k , or have a k th order contact) at a point x on a domain manifold M , denoted by $f_1 \sim f_2$, iff

$$\begin{aligned} f_1(x) &= f_2(x) && \text{called } 0\text{-tangent,} \\ \partial_x f_1(x) &= \partial_x f_2(x), && \text{called } 1\text{-tangent,} \\ \partial_{xx} f_1(x) &= \partial_{xx} f_2(x), && \text{called } 2\text{-tangent,} \\ &\dots && \text{etc. to the order } k. \end{aligned}$$

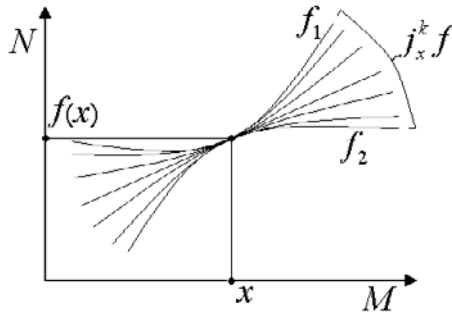


Fig. 5.1 An intuitive geometrical picture behind the k -jet concept, based on the idea of higher-order tangency or contact (see text for explanation).

In this way defined k -tangency is an equivalence relation, i.e.,

$$f_1 \sim f_2 \Rightarrow f_2 \sim f_1, \quad f_1 \sim f_2 \sim f_3 \Rightarrow f_1 \sim f_3, \quad f_1 \sim f_1.$$

Now a k -jet (or, a jet of order k), denoted by $j_x^k f$, of a smooth map $f : M \rightarrow N$ at a point $x \in M$ (see Figure 5.1), is defined as an equivalence class of k -tangent maps at x ,

$$j_x^k f = \{f' : f' \text{ is } k\text{-tangent to } f \text{ at } x\}.$$

The point x is called the *source* and the point $f(x)$ is the *target* of the k -jet $j_x^k f$.

We choose local coordinates on M and N in the neighborhood of the points x and $f(x)$, respectively. Then the k -jet $j_x^k f$ of any map close to f , at any point close to x , can be given by its Taylor-series expansion at

x , with coefficients up to degree k . Therefore, in a fixed coordinate chart, a k -jet can be identified with the collection of Taylor coefficients up to degree k .

The set of all k -jets of smooth maps from M to N is called the k -jet space and denoted by $J^k(M, N)$. It has a natural smooth manifold structure. Also, a map from a k -jet space $J^k(M, N)$ to a smooth manifold M or N is called a *jet bundle* (we will make this notion more precise later).

For example, consider a simple function $f : X \rightarrow Y, x \mapsto y = f(x)$, mapping the X -axis into the Y -axis. In this case, $M = X$ is a domain and $N = Y$ is a codomain. A 0-jet at a point $x \in X$ is given by its graph $(x, f(x))$. A 1-jet is given by a triple $(x, f(x), f'(x))$, a 2-jet is given by a quadruple $(x, f(x), f'(x), f''(x))$, and so on up to the order k (where $f'(x) = \frac{df(x)}{dx}$, etc.). The set of all k -jets from X to Y is called the k -jet space $J^k(X, Y)$.

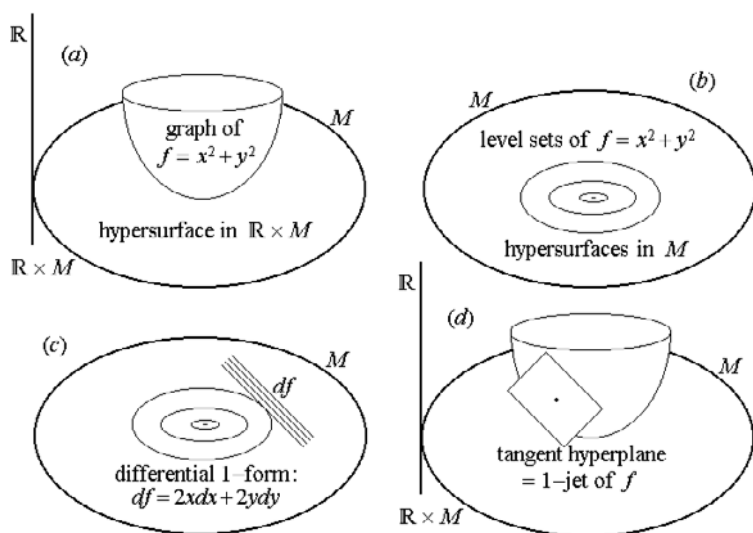


Fig. 5.2 Common spaces associated with a function f on a smooth manifold M (see text for explanation).

In case of a function of two variables, $f(x, y)$, the common spaces related to f , including its 1-jet $j^1 f$, are depicted in Figure 5.2 (see [Omohundro (1986)]). Recall that a *hypersurface* is a *codimension-1 submanifold*. Given a sample function $f(x, y) = x^2 + y^2$ in $M = \mathbb{R}^2$, then: (a) shows its graph as a hypersurface in $\mathbb{R} \times M$; (b) shows its level sets in M ; (c) shows its

differential form $df = 2xdx + 2ydy$ in the cotangent bundle T^*M ; and (d) shows a tangent hyperplane at a point $(x_0, y_0) \in M$ to its graph in $\mathbb{R} \times M$, which is a 1-jet $j_{(x_0, y_0)}^1 f$ to f at (x_0, y_0) . Note that $j_{(x_0, y_0)}^1 f$ is parallel to df , which means that its 1-jet space $J^1(\mathbb{R}, M)$ is an $(n+1)$ D extension of the cotangent bundle T^*M .

In mechanics we will consider a pair of maps $f_1, f_2 : \mathbb{R} \rightarrow M$ from the real line \mathbb{R} , representing the *time* t -axis, into a smooth n D (configuration) manifold M . We say that the maps $f_1 = f_1(t)$ and $f_2 = f_2(t)$ have the same k -jet $j_x^k f$ at a specified time instant $t_0 \in \mathbb{R}$, iff:

- (1) $f_1(t) = f_2(t)$ at $t_0 \in \mathbb{R}$, and also
- (2) the first k terms of their Taylor-series expansion around $t_0 \in \mathbb{R}$ are equal.

The k -jet space $J^k(\mathbb{R}, M)$ is the set of all k -jets $j_x^k f$ from \mathbb{R} to M .

Now, the fundamental geometrical construct in *time-dependent mechanics* is its *configuration fibre bundle* (see section 5.6 below). Given a configuration fibre bundle $M \rightarrow \mathbb{R}$ over the time axis \mathbb{R} , we say that the 1-jet space $J^1(\mathbb{R}, M)$ is the set of equivalence classes $j_t^1 s$ of sections $s^i : \mathbb{R} \rightarrow M$ of the bundle $M \rightarrow \mathbb{R}$, which are identified by their values $s^i(t)$, and by the values of their partial derivatives $\partial_t s^i = \partial_t s^i(t)$ at time points $t \in \mathbb{R}$. The 1-jet space $J^1(\mathbb{R}, M)$ is coordinated by (t, x^i, \dot{x}^i) , so the 1-jets are local coordinate maps

$$j_t^1 s : t \mapsto (t, x^i, \dot{x}^i).$$

Similarly, the 2-jet space $J^2(\mathbb{R}, M)$ is the set of equivalence classes $j_t^2 s$ of sections $s^i : \mathbb{R} \rightarrow M$ of the bundle $M \rightarrow \mathbb{R}$, which are identified by their values $s^i(t)$, as well as the values of their first and second partial derivatives, $\partial_t s^i$ and $\partial_{tt} s^i$, at time points $t \in \mathbb{R}$. The 2-jet space $J^2(\mathbb{R}, M)$ is coordinated by $(t, x^i, \dot{x}^i, \ddot{x}^i)$, so the 2-jets are local coordinate maps

$$j_t^2 s : t \mapsto (t, x^i, \dot{x}^i, \ddot{x}^i).$$

Generalization to the k -jet space $J^k(\mathbb{R}, M)$ is obvious. This mechanical jet formalism will be developed in section 5.6 below.

More generally, in a *physical field* context, instead of the mechanical configuration bundle over the time axis \mathbb{R} , we have some general physical fibre bundle $Y \rightarrow X$ over some smooth manifold (base) X . In this general context, the k -jet space $J^k(X, Y)$ of a bundle $Y \rightarrow X$ is the set of equivalence classes $j_x^k s$ of sections $s^i : X \rightarrow Y$, which are identified by their values

$s^i(x)$, as well as the first k terms of their Taylor-series expansion at points $x \in X$. This has two important physical consequences:

- (1) The k -jet space of sections $s^i : X \rightarrow Y$ of a fibre bundle $Y \rightarrow X$ is itself an nD smooth manifold, and
- (2) A k th-order differential operator on sections $s^i(x)$ of a fibre bundle $Y \rightarrow X$ can be described as a map of $J^k(X, Y)$ to a vector bundle over the base X .

A map from a k -jet space $J^k(X, Y)$ to a smooth manifold Y or X is called a *jet bundle*.

As a consequence, the dynamics of mechanical and physical field systems is played out on nD configuration and phase manifolds. Moreover, this dynamics can be phrased in geometrical terms due to the 1-1 correspondence between sections of the jet bundle $J^1(X, Y) \rightarrow Y$ and connections on the fibre bundle $Y \rightarrow X$.

In the framework of the standard *first-order Lagrangian formalism*, the nD *configuration space of sections* $s^i : X \rightarrow Y$ of a fibre bundle $Y \rightarrow X$ is the 1-jet space $J^1(X, Y)$, coordinated by $(x^\alpha, y^i, y_\alpha^i)$, where (x^α, y^i) are fibre coordinates of Y , while y_α^i are the so-called ‘derivative coordinates’ or ‘velocities’. A first-order *Lagrangian density*¹ on the *configuration manifold* $J^1(X, Y)$ is given by an exterior one-form (the so-called *horizontal density*)

$$L = \mathcal{L}(x^\alpha, y^i, y_\alpha^i)\omega, \quad \text{with} \quad \omega = dx^1 \wedge \dots \wedge dx^n.$$

This physical jet formalism will be developed below.

5.2 Definition of a 1-Jet Space

As introduced above, a 1-jet is defined as an equivalence class of functions having the same value and the same first derivatives at some designated point of the domain manifold (see Figure 5.1). Recall that in mechanical

¹Recall that in classical field theory, a distinction is made between the Lagrangian L , of which the *action* is the time integral $S[x^i] = \int L[x^i, \dot{x}^i]dt$ and the Lagrangian density \mathcal{L} , which one integrates over all space-time to get the action $S[\varphi^k] = \int \mathcal{L}[\varphi^k[x^i]]d^4x$. The Lagrangian is then the spatial integral of the Lagrangian density. However, \mathcal{L} is also frequently simply called the Lagrangian, especially in modern use; it is far more useful in relativistic theories since it is a locally defined, Lorentz scalar field. Both definitions of the Lagrangian can be seen as special cases of the general form, depending on whether the spatial variable x^i is incorporated into the index i or the parameters s in $\varphi^k[x^i]$. Quantum field theories are usually described in terms of \mathcal{L} , and the terms in this form of the Lagrangian translate quickly to the rules used in evaluating Feynman diagrams.

settings, the 1-jets are local coordinate maps

$$j_t^1 s : t \mapsto (t, x^i, \dot{x}^i).$$

More generally, given a fibre bundle $Y \rightarrow X$ with bundle coordinates (x^α, y^i) , consider the equivalence classes $j_x^1 s$ of its sections $s^i : X \rightarrow Y$, which are identified by their values $s^i(x)$ and the values of their first-order derivatives $\partial_\alpha s^i = \partial_\alpha s^i(x)$ at a point x on the domain (base) manifold X . They are called the 1-jets of sections s^i at $x \in X$. One can justify that the definition of jets is coordinate-independent by observing that the set $J^1(X, Y)$ of 1-jets $j_x^1 s$ is a smooth manifold with respect to the adapted coordinates $(x^\alpha, y^i, y_\alpha^i)$, such that [Sardanashvily (1993); Sardanashvily (1995); Giachetta *et. al.* (1997); Mangiarotti and Sardanashvily (2000a); Sardanashvily (2002a)]

$$y_\alpha^i(j_x^1 s) = \partial_\alpha s^i(x), \quad y_\alpha^i = \frac{\partial x^\mu}{\partial x'^\alpha} (\partial_\mu + y_\mu^j \partial_j) y'^i.$$

$J^1(X, Y)$ is called the 1-jet space of the fibre bundle $Y \rightarrow X$.

In other words, the 1-jets $j_x^1 s : x^\alpha \mapsto (x^\alpha, y^i, y_\alpha^i)$, which are first-order equivalence classes of sections of the fibre bundle $Y \rightarrow X$, can be identified with their codomain set of adapted coordinates on $J^1(X, Y)$,

$$j_x^1 s \equiv (x^\alpha, y^i, y_\alpha^i).$$

Note that in a section 5.7 below, the mechanical 1-jet space $J^1(\mathbb{R}, M) \equiv \mathbb{R} \times TM$ will be regarded as a fibre bundle over the base product-manifold $\mathbb{R} \times M$ (see [Neagu and Udrişte (2000a); Udrişte (2000); Neagu (2002); Neagu and Udrişte (2000b); Neagu (2000)] for technical details).

The jet space $J^1(X, Y)$ admits the natural fibrations

$$\pi^1 : J^1(X, Y) \ni j_x^1 s \mapsto x \in X, \quad \text{and} \quad (5.1)$$

$$\pi_0^1 : J^1(X, Y) \ni j_x^1 s \mapsto s(x) \in Y, \quad (5.2)$$

which form the commutative triangle:

$$\begin{array}{ccc} J^1(X, Y) & \xrightarrow{\pi_0^1} & Y \\ & \searrow \pi^1 & \swarrow \pi \\ & X & \end{array}$$

It is convenient to call π^1 (5.1) the *jet bundle*, while π_0^1 (5.2) is called the *affine jet bundle*. Note that, if $Y \rightarrow X$ is a vector or an affine bundle, it also holds for the jet bundle π_1 (5.1) [Sardanashvily (1993); Sardanashvily (1995); Giachetta *et. al.* (1997); Mangiarotti and Sardanashvily (2000a); Sardanashvily (2002a)].

There exist several equivalent ways in order to give the 1-jet space $J^1(X, Y)$ with the smooth manifold structure. Let $Y \rightarrow X$ be a fibre bundle with fibred coordinate atlases (4.4). The 1-jet space $J^1(X, Y)$ of the bundle $Y \rightarrow X$ admits the adapted coordinate atlases

$$(x^\alpha, y^i, y_\alpha^i), \quad (x^\alpha, y^i, y_\alpha^i)(j_x^1 s) = (x^\alpha, s^i(x), \partial_\alpha s^i), \quad (5.3)$$

$$y_\alpha'^i = \left(\frac{\partial y'^i}{\partial y^j} y_\mu^j + \frac{\partial y'^i}{\partial x^\mu} \right) \frac{\partial x^\mu}{\partial x'^\alpha}, \quad (5.4)$$

and thus satisfies the conditions which are required of a manifold. The surjection (5.1) is a bundle. The surjection (5.2) is a bundle. If $Y \rightarrow X$ is a bundle, so is the surjection (5.1).

The transformation law (5.4) shows that the jet bundle $J^1(X, Y) \rightarrow Y$ is an affine bundle. It is modelled on the vector bundle $T^*X \otimes VY \rightarrow Y$. In particular, if Y is the trivial bundle

$$\pi_2 : V \times \mathbb{R}^m \rightarrow \mathbb{R}^m,$$

the corresponding jet bundle $J^1(X, Y) \rightarrow \mathbb{R}^m$ (5.1) is a trivial bundle.

There exist the following two canonical bundle monomorphisms of the jet bundle $J^1(X, Y) \rightarrow Y$ [Sardanashvily (1993); Sardanashvily (1995); Giachetta *et. al.* (1997); Mangiarotti and Sardanashvily (2000a); Sardanashvily (2002a)]:

- the *contact map*

$$\lambda : J^1(X, Y) \hookrightarrow T^*X \otimes TY, \quad \lambda = dx^\alpha \otimes \widehat{\partial}_\alpha = dx^\alpha \otimes (\partial_\alpha + y_\alpha^i \partial_i), \quad (5.5)$$

- the *complementary map*

$$\theta : J^1(X, Y) \hookrightarrow T^*Y \otimes VY, \quad \theta = \widehat{dy}^i \otimes \partial_i = (dy^i - y_\alpha^i dx^\alpha) \otimes \partial_i. \quad (5.6)$$

These canonical maps enable us to express the jet-space machinery in terms of tangent-valued differential forms (see section 4.10 above).

The operators

$$\widehat{\partial}_\alpha = \partial_\alpha + y_\alpha^i \partial_i$$

are usually called the *total derivatives*, or the *formal derivatives*, while the forms

$$\widehat{dy}^i = dy^i - y_\alpha^i dx^\alpha$$

are conventionally called the *contact forms*.

Identifying the 1-jet space $J^1(X, Y)$ to its images under the canonical maps (5.5) and (5.6), one can represent 1-jets $j_x^1 s \equiv (x^\alpha, y^i, y_\alpha^i)$ by tangent-valued forms

$$dx^\alpha \otimes (\partial_\alpha + y_\alpha^i \partial_i), \quad \text{and} \quad (dy^i - y_\alpha^i dx^\alpha) \otimes \partial_i. \quad (5.7)$$

There exists a *jet functor* $\mathfrak{J} : Bun \rightarrow Jet$, from the category *Bun* of fibre bundles to the category *Jet* of jet spaces. It implies the natural prolongation of maps of bundles to maps of jet spaces.

Every bundle map $\Phi : Y \rightarrow Y'$ over a diffeomorphism f of X has the 1-jet prolongation to the bundle map $j^1 \Phi : J^1(X, Y) \rightarrow J^1(X, Y)'$, given by

$$\begin{aligned} j^1 \Phi : j_{xs}^1 &\mapsto j_{f(x)}^1 (\Phi \circ s \circ f^{-1}), \\ y'^i_\alpha \circ j^1 \Phi &= \partial_\alpha (\Phi^i \circ f^{-1}) + \partial_j (\Phi^i y_\alpha^j \circ f^{-1}). \end{aligned} \quad (5.8)$$

It is both an affine bundle map over Φ and a fibred map over the diffeomorphism f . The 1-jet prolongations (5.8) of fibred maps satisfy the chain rules

$$j^1(\Phi \circ \Phi') = j^1 \Phi \circ j^1 \Phi', \quad j^1(\text{Id}_Y) = \text{Id}_{J^1(X, Y)}.$$

If Φ is a surjection (resp. an injection), so is $j^1 \Phi$.

In particular, every section s of a bundle $Y \rightarrow X$ admits the 1-jet prolongation to the section $j_x^1 s$ of the jet bundle $J^1(X, Y) \rightarrow X$, given by

$$(y^i, y_\alpha^i) \circ j_x^1 s = (s^i(x), \partial_\alpha s^i).$$

We have

$$\lambda \circ j_x^1 s = Ts,$$

where λ is the contact map (5.5).

Every projectable vector-field u on a fibre bundle $Y \rightarrow X$,

$$u = u^\alpha(x) \partial_\alpha + u^i(y) \partial_i$$

has the 1-jet lift to the projectable vector-field j^1u on the 1-jet space $J^1(X, Y)$, given by

$$\begin{aligned} j^1u &\equiv \bar{u} = r_1 \circ j^1u : J^1(X, Y) \rightarrow TJ^1(X, Y), \\ j^1u &\equiv \bar{u} = u^\alpha \partial_\alpha + u^i \partial_i + (d_\alpha u^i - y_\mu^i \partial_\alpha u^\mu) \partial_i^\alpha. \end{aligned} \quad (5.9)$$

Geometrical applications of jet spaces are based on the canonical map over $J^1(X, Y)$,

$$J^1(X, Y) \times TX \rightarrow J^1(X, Y) \times TY,$$

which means the canonical horizontal splitting of the tangent bundle TY determined over $J^1(X, Y)$ as follows [Sardanashvily (1993); Sardanashvily (1995); Giachetta *et. al.* (1997); Mangiarotti and Sardanashvily (2000a); Sardanashvily (2002a)].

The canonical maps (5.5) and (5.6) induce the bundle monomorphisms

$$\begin{aligned} \hat{\lambda} : J^1(X, Y) \times TX &\rightarrow J^1(X, Y) \times TY, & \partial_\alpha &\mapsto \hat{\partial}_\alpha = \partial_\alpha \rfloor \lambda \\ \hat{\theta} : J^1(X, Y) \times V^*Y &\rightarrow J^1(X, Y) \times T^*Y, & dy^i &\mapsto \hat{dy}^i = \theta \rfloor dy^i \end{aligned} \quad (5.10)$$

The map (5.10) determines the canonical horizontal splitting of the pull-back

$$\begin{aligned} J^1(X, Y) \times TY &= \hat{\lambda}(TX) \oplus VY, \\ \dot{x}^\alpha \partial_\alpha + \dot{y}^i \partial_i &= \dot{x}^\alpha (\partial_\alpha + y_\alpha^i \partial_i) + (\dot{y}^i - \dot{x}^\alpha y_\alpha^i) \partial_i. \end{aligned} \quad (5.12)$$

Similarly, the map (5.11) induces the dual canonical horizontal splitting of the pull-back

$$\begin{aligned} J^1(X, Y) \times T^*Y &= T^*X \oplus \hat{\theta}(V^*Y), \\ \dot{x}_\alpha dx^\alpha + \dot{y}_i dy^i &= (\dot{x}_\alpha + \dot{y}_i y_\alpha^i) dx^\alpha + \dot{y}_i (dy^i - y_\alpha^i dx^\alpha). \end{aligned} \quad (5.13)$$

Building on the canonical splittings (5.12) and (5.13), one gets the following canonical horizontal splittings of

- a projectable vector-field on a fibre bundle $Y \rightarrow X$,

$$u = u^\alpha \partial_\alpha + u^i \partial_i = u_H + u_V = u^\alpha (\partial_\alpha + y_\alpha^i \partial_i) + (u^i - u^\alpha y_\alpha^i) \partial_i, \quad (5.14)$$

- an exterior 1-form

$$\sigma = \sigma_\alpha dx^\alpha + \sigma_i dy^i = (\sigma_\alpha + y_\alpha^i \sigma_i) dx^\alpha + \sigma_i (dy^i - y_\alpha^i dx^\alpha),$$

- a tangent-valued projectable horizontal form

$$\begin{aligned}\phi &= dx^{\alpha_1} \wedge \cdots \wedge dx^{\alpha_r} \otimes (\phi_{\alpha_1 \dots \alpha_r}^\mu \partial_\mu + \phi_{\alpha_1 \dots \alpha_r}^i \partial_i) \\ &= dx^{\alpha_1} \wedge \cdots \wedge dx^{\alpha_r} \otimes [\phi_{\alpha_1 \dots \alpha_r}^\mu (\partial_\mu + y_\mu^i \partial_i) + (\phi_{\alpha_1 \dots \alpha_r}^i - \phi_{\alpha_1 \dots \alpha_r}^\mu y_\mu^i) \partial_i]\end{aligned}$$

and, e.g., the canonical 1-form

$$\begin{aligned}\theta_Y &= dx^\alpha \otimes \partial_\alpha + dy^i \otimes \partial_i = \alpha + \theta = dx^\alpha \otimes \widehat{\partial}_\alpha + \widehat{dy}^i \otimes \partial_i \\ &= dx^\alpha \otimes (\partial_\alpha + y_\alpha^i \partial_i) + (dy^i - y_\alpha^i dx^\alpha) \otimes \partial_i.\end{aligned}\quad (5.15)$$

The splitting (5.15) implies the canonical horizontal splitting of the exterior differential

$$d = d_{\theta_Y} = d_H + d_V = d_\alpha + d_\theta. \quad (5.16)$$

Its components d_H and d_V act on the pull-backs

$$\phi_{\alpha_1 \dots \alpha_r}(y) dx^{\alpha_1} \wedge \cdots \wedge dx^{\alpha_r}$$

of horizontal exterior forms on a bundle $Y \rightarrow X$ onto $J^1(X, Y)$ by π_{01} . In this case, d_H makes the sense of the total differential

$$d_H \phi_{\alpha_1 \dots \alpha_r}(y) dx^{\alpha_1} \wedge \cdots \wedge dx^{\alpha_r} = (\partial_\mu + y_\mu^i \partial_i) \phi_{\alpha_1 \dots \alpha_r}(y) dx^\mu \wedge dx^{\alpha_1} \wedge \cdots \wedge dx^{\alpha_r},$$

whereas d_V is the vertical differential

$$d_V \phi_{\alpha_1 \dots \alpha_r}(y) dx^{\alpha_1} \wedge \cdots \wedge dx^{\alpha_r} = \partial_i \phi_{\alpha_1 \dots \alpha_r}(y) (dy^i - y_\mu^i dx^\mu) \wedge dx^{\alpha_1} \wedge \cdots \wedge dx^{\alpha_r}.$$

If $\phi = \widetilde{\phi} \omega$ is an exterior horizontal density on $Y \rightarrow X$, we have

$$d\phi = d_V \phi = \partial_i \widetilde{\phi} dy^i \wedge \omega.$$

5.3 Connections as Jet Fields

Recall that one can introduce the notion of *connections* on a general fibre bundle $Y \rightarrow X$ in several equivalent ways. In this section, following [Giachetta *et al.* (1997); Kolar *et al.* (1993); Mangiarotti and Sardanashvily (2000a); Saunders (1989)], we start from the traditional definition of a connection as a *horizontal splitting* of the tangent space to Y at every point $y \in Y$.

A connection on a fibre bundle $Y \rightarrow X$ is usually defined as a linear bundle monomorphism

$$\Gamma : Y \times TX \rightarrow TY, \quad \Gamma : \dot{x}^\alpha \partial_\alpha \mapsto \dot{x}^\alpha (\partial_\alpha + \Gamma_\alpha^i(y) \partial_i), \quad (5.17)$$

which splits the exact sequence (4.13), i.e.,

$$\pi_T \circ \Gamma = \text{Id}_{Y \times TX}.$$

The image HY of $Y \times TX$ by a connection Γ is called the *horizontal distribution*. It splits the tangent bundle TY as

$$\begin{aligned} TY &= HY \oplus VY, & \text{giving} \\ \dot{x}^\alpha \partial_\alpha + \dot{y}^i \partial_i &= \dot{x}^\alpha (\partial_\alpha + \Gamma_\alpha^i \partial_i) + (\dot{y}^i - \dot{x}^\alpha \Gamma_\alpha^i) \partial_i. \end{aligned} \quad (5.18)$$

Similarly, horizontal splitting of the cotangent bundle,

$$\begin{aligned} T^*Y &= T^*X \oplus \Gamma(V^*X), & \text{gives} \\ \dot{x}_\alpha dx^\alpha + \dot{y}_i dy^i &= (\dot{x}_\alpha + \Gamma_\alpha^i \dot{y}_i) dx^\alpha + \dot{y}_i (dy^i - \Gamma_\alpha^i dx^\alpha). \end{aligned}$$

Alternatively, a connection on a fibre bundle $Y \rightarrow X$ can be defined as a *jet field*, i.e., a *section* of the affine jet bundle $J^1(X, Y) \rightarrow Y$. This connection is called the *Ehresmann connection*, and historically it was a primary reason for C. Ehresmann to develop the concept of jet spaces.

Due to the Theorem that says [Hirzebruch (1966)]: Every exact sequence of vector bundles (4.10) is split, a jet-field connection on a fibre bundle always exists.

A connection on a fibre bundle $Y \rightarrow X$ is defined to be a tangent-valued projectable horizontal one-form Γ on Y such that $\Gamma|_\phi = \phi$ for all exterior horizontal 1-forms ϕ on Y . It is given by the coordinate expression

$$\Gamma = dx^\alpha \otimes (\partial_\alpha + \Gamma_\alpha^i(y) \partial_i), \quad \Gamma_\alpha^i = \left(\frac{\partial y'^i}{\partial y^j} \Gamma_\mu^j + \frac{\partial y'^i}{\partial x^\mu} \right) \frac{\partial x^\mu}{\partial x'^\alpha}, \quad (5.19)$$

such that $\Gamma(\partial_\alpha) = \partial_\alpha|_\Gamma$. Conversely, every horizontal tangent-valued 1-form on a fibre bundle $Y \rightarrow X$ which projects onto the canonical tangent-valued form θ_X (4.137) on X defines a connection on $Y \rightarrow X$.

In an equivalent way, the horizontal splitting (5.18) is given by the vertical-valued form

$$\Gamma = (dy^i - \Gamma_\alpha^i dx^\alpha) \otimes \partial_i, \quad (5.20)$$

which determines the epimorphism

$$\Gamma : TY \rightarrow VY, \quad \dot{x}^\alpha \partial_\alpha + \dot{y}^i \partial_i \mapsto (\dot{x}^\alpha \partial_\alpha + \dot{y}^i \partial_i)|_\Gamma = (\dot{y}^i - \dot{x}^\alpha \Gamma_\alpha^i) \partial_i.$$

Let $Y \rightarrow X$ be a vector bundle. A *linear connection* on Y reads

$$\Gamma = dx^\alpha \otimes [\partial_\alpha - \Gamma_{j\alpha}^i(x) y^j \partial_i]. \quad (5.21)$$

Let $Y \rightarrow X$ be an affine bundle modelled on a vector bundle $\bar{Y} \rightarrow X$. An *affine connection* on Y reads

$$\begin{aligned}\Gamma &= dx^\alpha \otimes [\partial_\alpha + (-\Gamma^i_{j\alpha}(x)y^j + \Gamma^i_\alpha(x))\partial_i], \quad \text{where} \\ \bar{\Gamma} &= dx^\alpha \otimes [\partial_\alpha - \Gamma^i_{j\alpha}(x)\bar{y}^j\partial_i] \quad \text{is a linear connection on } \bar{Y}.\end{aligned}$$

Since the affine jet bundle $J^1(X, Y) \rightarrow Y$ is modelled on the vector bundle $Y \rightarrow X$, Ehresmann connections on $Y \rightarrow X$ constitute an affine space modelled on the linear space of soldering forms on Y . If Γ is a connection and σ is a soldering form (4.138) on Y , its sum

$$\Gamma + \sigma = dx^\alpha \otimes [\partial_\alpha + (\Gamma^i_\alpha + \sigma^i_\alpha)\partial_i]$$

is a connection on Y . Conversely, if Γ and Γ' are connections on Y , then

$$\Gamma - \Gamma' = (\Gamma^i_\alpha - \Gamma'^i_\alpha)dx^\alpha \otimes \partial_i$$

is a soldering form.

Given a connection Γ , a vector-field u on a fibre bundle $Y \rightarrow X$ is called horizontal if it lives in the horizontal distribution HY , i.e., takes the form

$$u = u^\alpha(y)(\partial_\alpha + \Gamma^i_\alpha(y)\partial_i). \quad (5.22)$$

Any vector-field τ on the base X of a fibre bundle $Y \rightarrow X$ admits the horizontal lift

$$\Gamma\tau = \tau \rfloor \Gamma = \tau^\alpha(\partial_\alpha + \Gamma^i_\alpha\partial_i) \quad (5.23)$$

onto Y by means of a connection Γ on $Y \rightarrow X$.

Given the splitting (5.17), the dual splitting of the exact sequence (4.14) is

$$\Gamma : V^*Y \rightarrow T^*Y, \quad \bar{d}y^i \mapsto \Gamma \rfloor \bar{d}y^i = dy^i - \Gamma^i_\alpha dx^\alpha, \quad (5.24)$$

where Γ is the vertical-valued form (5.20).

There is 1–1 correspondence between the connections on a fibre bundle $Y \rightarrow X$ and the jet fields, i.e., global sections of the affine jet bundle $J^1(X, Y) \rightarrow Y$. Indeed, given a global section Γ of $J^1(X, Y) \rightarrow Y$, the tangent-valued form

$$\lambda \circ \Gamma = dx^\alpha \otimes (\partial_\alpha + \Gamma^i_\alpha\partial_i)$$

gives the horizontal splitting (5.18) of TY . Therefore, the vertical-valued form

$$\theta \circ \Gamma = (dy^i - \Gamma_\alpha^i dx^\alpha) \otimes \partial_i$$

leads to the dual splitting (5.24).

It follows immediately from this definition that connections on a fibre bundle $Y \rightarrow X$ constitute an affine space modelled over the vector space of soldering forms σ (4.138). They obey the *coordinate transformation law* [Giachetta *et. al.* (1997); Saunders (1989)]

$$\Gamma_\alpha'^i = \frac{\partial x^\mu}{\partial x'^\alpha} (\partial_\mu + \Gamma_\mu^j \partial_j) y'^i.$$

In particular, a *linear connection* K on the tangent bundle TX of a manifold X and the *dual connection* K^* to K on the cotangent bundle T^*X are given by the coordinate expressions

$$K_\beta^\alpha = -K^\alpha_{\nu\beta}(x) \dot{x}^\nu, \quad K_{\alpha\beta}^* = K^\nu_{\alpha\beta}(x) \dot{x}_\nu. \quad (5.25)$$

Also, given a connection Γ on $Y \rightarrow X$, the vertical tangent map $V\Gamma : VY \rightarrow J^1(\mathbb{R}, V)Y$ induces the *vertical connection*

$$V\Gamma = dx^\alpha \otimes (\partial_\alpha + \Gamma_\alpha^i \partial_{y^i} + \partial_V \Gamma_\alpha^i \partial_{\dot{y}^i}), \quad \partial_V \Gamma_\alpha^i = \dot{y}^j \partial_j \Gamma_\alpha^i, \quad (5.26)$$

on the bundle $VY \rightarrow X$. The connection $V\Gamma$ is projectable to the connection Γ on Y , and it is a linear bundle map over Γ . The dual *coververtical connection* on the bundle $V^*Y \rightarrow X$ reads

$$V^*\Gamma = dx^\alpha \otimes (\partial_\alpha + \Gamma_\alpha^i \partial_{y^i} - \partial_j \Gamma_\alpha^i \dot{y}_i \partial_{\dot{y}^j}). \quad (5.27)$$

Connections on a bundle $Y \rightarrow X$ constitute the affine space modelled on the linear space of soldering 1-forms on Y . It means that, if Γ is a connection and σ is a soldering form on a bundle Y , its sum

$$\Gamma + \sigma = dx^\alpha \otimes [\partial_\alpha + (\Gamma_\alpha^i + \sigma_\alpha^i) \partial_i]$$

is a connection on Y . Conversely, if Γ and Γ' are connections on a bundle Y , then their difference

$$\Gamma - \Gamma' = (\Gamma_\alpha^i - \Gamma_\alpha'^i) dx^\alpha \otimes \partial_i$$

is a soldering form on Y .

Given a fibre bundle $Y \rightarrow X$, let $f : X' \rightarrow X$ be a manifold map and f^*Y the pull-back of Y over X' . Any connection Γ (5.20) on $Y \rightarrow X$ induces the pull-back connection

$$f^*\Gamma = (dy^i - \Gamma_\alpha^i(f^\mu(x')\nu), y^j) \frac{\partial f^\alpha}{\partial x'^\mu} dx'^\mu \otimes \partial_i \quad (5.28)$$

on the pull-back fibre bundle $f^*Y \rightarrow X'$.

Since the affine jet bundle $J^1(X, Y) \rightarrow Y$ is modelled on the vector bundle $Y \rightarrow X$, connections on a fibre bundle Y constitute the affine space modelled on the linear space of soldering forms on Y . It follows that, if Γ is a connection and

$$\sigma = \sigma_\alpha^i dx^\alpha \otimes \partial_i$$

is a soldering form on a fibre bundle Y , its sum

$$\Gamma + \sigma = dx^\alpha \otimes [\partial_\alpha + (\Gamma_\alpha^i + \sigma_\alpha^i) \partial_i]$$

is a connection on Y . Conversely, if Γ and Γ' are connections on a fibre bundle Y , then

$$\Gamma - \Gamma' = (\Gamma_\alpha^i - \Gamma'^i_\alpha) dx^\alpha \otimes \partial_i$$

is a soldering form on Y .

The key point for physical applications lies in the fact that every connection Γ on a fibre bundle $Y \rightarrow X$ induces the first-order differential operator

$$D^\Gamma : J^1(X, Y) \rightarrow T^*X \otimes VY, \quad D^\Gamma = \lambda - \Gamma \circ \pi_0^1 = (y_\alpha^i - \Gamma_\alpha^i) dx^\alpha \otimes \partial_i, \quad (5.29)$$

called the *covariant differential* relative to the connection Γ . If $s : X \rightarrow Y$ is a section, one defines its covariant differential

$$\nabla^\Gamma s = D_\Gamma \circ j^1 s = (\partial_\alpha s^i - \Gamma_\alpha^i \circ s) dx^\alpha \otimes \partial_i \quad (5.30)$$

and its *covariant derivative*

$$\nabla_\tau^\Gamma s = \tau \rfloor \nabla^\Gamma s \quad (5.31)$$

along a vector-field τ on X . A (local) section s of $Y \rightarrow X$ is said to be an *integral section* of a connection Γ (or parallel with respect to Γ) if s obeys the equivalent conditions

$$\nabla^\Gamma s = 0 \quad \text{or} \quad j^1 s = \Gamma \circ s. \quad (5.32)$$

Furthermore, if $s : X \rightarrow Y$ is a global section, there exists a connection Γ such that s is an integral section of Γ . This connection is defined as an extension of the local section $s(x) \mapsto j^1 s(x)$ of the affine jet bundle $J^1(X, Y) \rightarrow Y$ over the closed imbedded submanifold $s(X) \subset Y$.

Note that every connection Γ on the bundle $Y \rightarrow X$ defines a system of first-order differential equations on Y (in the spirit of [Bryant *et. al.* (1991); Krasil'shchik *et. al.* (1985); Pommaret (1978)]) which is an imbedded sub-bundle $\Gamma(Y) = \text{Ker } D_\Gamma$ of the jet bundle $J^1(X, Y) \rightarrow Y$. It is given by the coordinate relations

$$y_\alpha^i = \Gamma^i(y). \quad (5.33)$$

Integral sections for Γ are local solutions of (5.33), and *vice versa*.

We can introduce the following basic forms involving a connection Γ and a soldering form σ :

- the *curvature of a connection* Γ is given by the horizontal vertical-valued two-form:

$$\begin{aligned} R &= \frac{1}{2} d_\Gamma \Gamma = \frac{1}{2} R_{\alpha\mu}^i dx^\alpha \wedge dx^\mu \otimes \partial_i, \\ R_{\alpha\mu}^i &= \partial_\alpha \Gamma_\mu^i - \partial_\mu \Gamma_\alpha^i + \Gamma_\alpha^j \partial_j \Gamma_\mu^i - \Gamma_\mu^j \partial_j \Gamma_\alpha^i; \end{aligned} \quad (5.34)$$

- the *torsion of a connection* Γ with respect to σ :

$$\begin{aligned} \Omega &= d_\sigma \Gamma = d_\Gamma \sigma = \frac{1}{2} \Omega_{\alpha\mu}^i dx^\alpha \wedge dx^\mu \otimes \partial_i \\ &= (\partial_\alpha \sigma_\mu^i + \Gamma_\alpha^j \partial_j \sigma_\mu^i - \partial_j \Gamma_\alpha^i \sigma_\mu^j) dx^\alpha \wedge dx^\mu \otimes \partial_i; \end{aligned} \quad (5.35)$$

- the *soldering curvature* of σ :

$$\begin{aligned} \varepsilon &= \frac{1}{2} d_\sigma \sigma = \frac{1}{2} \varepsilon_{\alpha\mu}^i dx^\alpha \wedge dx^\mu \otimes \partial_i \\ &= \frac{1}{2} (\sigma_\alpha^j \partial_j \sigma_\mu^i - \sigma_\mu^j \partial_j \sigma_\alpha^i) dx^\alpha \wedge dx^\mu \otimes \partial_i. \end{aligned} \quad (5.36)$$

They satisfy the following relations:

$$\Gamma' = \Gamma + \sigma, \quad R' = R + \varepsilon + \Omega, \quad \Omega' = \Omega + 2\varepsilon.$$

In particular, the curvature (5.34) of the linear connection (5.21) reads

$$\begin{aligned} R_{\alpha\mu}^i(y) &= -R_{j\alpha\mu}^i(x)y^j, \\ R_{j\alpha\mu}^i &= \partial_\alpha \Gamma_{j\mu}^i - \partial_\mu \Gamma_{j\alpha}^i + \Gamma_{j\mu}^k \Gamma_{k\alpha}^i - \Gamma_{j\alpha}^k \Gamma_{k\mu}^i. \end{aligned}$$

Let Y and Y' be vector bundles over X . Given linear connections Γ and Γ' on Y and Y' respectively, there is the unique linear connection $\Gamma \otimes \Gamma'$ on the tensor product $Y \otimes Y' \rightarrow X$, such that the following diagram commutes:

$$\begin{array}{ccc} J^1(X, Y) \times J^1(X, Y)' & \xrightarrow{J^1 \otimes} & J^1(Y \otimes Y') \\ \Gamma \times \Gamma' \downarrow & & \downarrow \Gamma \otimes \Gamma' \\ Y \times Y' & \xrightarrow{\otimes} & Y \otimes Y' \end{array}$$

It is called the *tensor-product connection* and has the coordinate expression

$$(\Gamma \otimes \Gamma')_{\alpha}^{ik} = \Gamma_{j\alpha}^i y^{jk} + \Gamma'_{j\alpha}^k y^{ij}.$$

Every connection Γ on $Y \rightarrow X$, by definition, induces the *horizontal distribution* on Y ,

$$\Gamma : TX \hookrightarrow TY, \quad \text{locally given by} \quad \partial_{\alpha} \mapsto \partial_{\alpha} + \Gamma_{\alpha}^i(y) \partial_i.$$

It is generated by horizontal lifts

$$\tau_{\Gamma} = \tau^{\alpha} (\partial_{\alpha} + \Gamma_{\alpha}^i \partial_i)$$

onto Y of vector-fields $\tau = \tau^{\alpha} \partial_{\alpha}$ on X . The associated Pfaffian system is locally generated by the forms $(dy^i - \Gamma_{\alpha}^i dx^{\alpha})$.

The horizontal distribution $\Gamma(TX)$ is involutive iff Γ is a *curvature-free connection*. As a proof, straightforward calculations show that $[\tau_{\Gamma}, \tau'_{\Gamma}] = ([\tau, \tau'])_{\Gamma}$ iff the curvature R (5.34) of Γ vanishes everywhere.

Not every bundle admits a curvature-free connection. If a principal bundle over a simply-connected base (i.e., its first homotopy group is trivial) admits a curvature-free connection, this bundle is trivializable [Kobayashi and Nomizu (1963/9)].

The horizontal distribution defined by a curvature-free connection is completely integrable. The corresponding foliation on Y is transversal to the foliation defined by the fibration $\pi : Y \rightarrow X$. It is called the *horizontal foliation*. Its leaf through a point $y \in Y$ is defined locally by the integral section s_y of the connection Γ through y . Conversely, let Y admits a horizontal foliation such that, for each point $y \in Y$, the leaf of this foliation through y is locally defined by some section s_y of $Y \rightarrow X$ through y . Then, the following map is well defined

$$\Gamma : Y \rightarrow J^1(X, Y), \quad \Gamma(y) = j_s^1 s_y, \quad \pi(y) = x.$$

This is a curvature-free connection on Y . There is the 1-1 correspondence between the curvature-free connections and the horizontal foliations on a bundle $Y \rightarrow X$.

Given a horizontal foliation on $Y \rightarrow X$, there exists the associated atlas of bundle coordinates (x^α, y^i) of Y such that (i) every leaf of this foliation is local generated by the equations $y^i = \text{const}$, and (ii) the transition functions $y^i \rightarrow y'^i(y^j)$ are independent on the coordinates x^α of the base X [Kamber and Tondeur (1975)]. It is called the *atlas of constant local trivializations*. Two such atlases are said to be equivalent if their union also is an atlas of constant local trivializations. They are associated with the same horizontal foliation.

There is the 1-1 correspondence between the curvature-free connections Γ on a bundle $Y \rightarrow X$ and the equivalence classes of atlases Ψ_c of constant local trivializations of Y such that $\Gamma_\alpha^i = 0$ relative to the coordinates of the corresponding atlas Ψ_c [Canarutto (1986)].

Connections on a bundle over a 1D base X^1 are curvature-free connections.

In particular, let $Y \rightarrow X^1$ be such a bundle ($X^1 = \mathbb{R}$ or $X^1 = S^1$). It is coordinated by (t, y^i) , where t is either the canonical parameter of \mathbb{R} or the standard local coordinate of S^1 together with the transition functions $t' = t + \text{const}$. Relative to this coordinate, the base X^1 admits the standard vector-field ∂_t and the standard one-form dt . Let Γ be a connection on $Y \rightarrow X^1$. Such a connection defines a horizontal foliation on $Y \rightarrow X^1$. Its leaves are the integral curves of the horizontal lift

$$\tau_\Gamma = \partial_t + \Gamma^i \partial_i \quad (5.37)$$

of ∂_t by Γ . The corresponding Pfaffian system is locally generated by the forms $(dy^i - \Gamma^i dt)$. There exists an atlas of constant local trivializations (t, y^i) such that $\Gamma^i = 0$ and $\tau_\Gamma = \partial_t$ relative to these coordinates.

A connection Γ on $Y \rightarrow X^1$ is called *complete* if the horizontal vector-field (5.37) is complete. Every trivialization of $Y \rightarrow \mathbb{R}$ defines a complete connection. Conversely, every complete connection on $Y \rightarrow \mathbb{R}$ defines a trivialization $Y \simeq \mathbb{R} \times M$. The vector-field (5.37) becomes the vector-field ∂_t on $\mathbb{R} \times M$. As a proof, every trivialization of $Y \rightarrow \mathbb{R}$ defines a 1-parameter group of isomorphisms of $Y \rightarrow \mathbb{R}$ over $\text{Id}_\mathbb{R}$, and hence a complete connection. Conversely, let Γ be a complete connection on $Y \rightarrow \mathbb{R}$. The vector-field τ_Γ (5.37) is the generator of a 1-parameter group G_Γ which acts freely on Y . The orbits of this action are the integral sections of τ_Γ . Hence we get a projection $Y \rightarrow M = Y/G_\Gamma$ which, together with the projection

$Y \rightarrow \mathbb{R}$, defines a trivialization $Y \simeq \mathbb{R} \times M$.

Let us consider a bundle $\pi : Y \rightarrow X$ which admits a *composite fibration*

$$Y \rightarrow \Sigma \rightarrow X, \quad (5.38)$$

where $Y \rightarrow \Sigma$ and $\Sigma \rightarrow X$ are bundles. It is equipped with the bundle coordinates $(x^\alpha, \sigma^m, y^i)$ together with the transition functions

$$x^\alpha \rightarrow x'^\alpha(x^\mu), \quad \sigma^m \rightarrow \sigma'^m(x^\mu, \sigma^n), \quad y^i \rightarrow y'^i(x^\mu, \sigma^n, y^j),$$

where (x^μ, σ^m) are bundle coordinates of $\Sigma \rightarrow X$. For example, we have the composite bundles

$$TY \rightarrow Y \rightarrow X, \quad VY \rightarrow Y \rightarrow X, \quad J^1(X, Y) \rightarrow Y \rightarrow X.$$

Let

$$A = dx^\alpha \otimes (\partial_\alpha + A_\alpha^i \partial_i) + d\sigma^m \otimes (\partial_m + A_m^i \partial_i) \quad (5.39)$$

be a connection on the bundle $Y \rightarrow \Sigma$ and

$$\Gamma = dx^\alpha \otimes (\partial_\alpha + \Gamma_\alpha^m \partial_m)$$

a connection on the bundle $\Sigma \rightarrow X$. Given a vector-field τ on X , let us consider its horizontal lift τ_Γ onto Σ by Γ and then the horizontal lift $(\tau_\Gamma)_A$ of τ_Γ onto Y by the connection (5.39).

There exists the connection

$$\gamma = dx^\alpha \otimes [\partial_\alpha + \Gamma_\alpha^m \partial_m + (A_m^i \Gamma_\alpha^m + A_\alpha^i) \partial_i]. \quad (5.40)$$

on $Y \rightarrow X$ such that the horizontal lift τ_γ onto Y of any vector-field τ on X consists with the above lift $(\tau_\Gamma)_A$ [Sardanashvily (1993); Sardanashvily (1995)]. It is called the *composite connection*.

Given a composite bundle Y (5.38), the exact sequence

$$0 \rightarrow VY_\Sigma \hookrightarrow VY \rightarrow Y \times V\Sigma \rightarrow 0$$

over Y take place, where VY_Σ is the vertical tangent bundle of $Y \rightarrow \Sigma$. Every connection (5.39) on the bundle $Y \rightarrow \Sigma$ induces the splitting

$$VY = VY_\Sigma \oplus (Y \times V\Sigma), \quad \text{given by} \\ \dot{y}^i \partial_i + \dot{\sigma}^m \partial_m = (\dot{y}^i - A_m^i \dot{\sigma}^m) \partial_i + \dot{\sigma}^m (\partial_m + A_m^i \partial_i).$$

Due to this splitting, one can construct the first-order differential operator

$$\begin{aligned}\tilde{D} &= \pi_1 \circ D_\gamma : J^1(X, Y) \rightarrow T^*X \otimes VY \rightarrow T^*X \otimes VY_\Sigma, \\ \tilde{D} &= dx^\alpha \otimes (y_\alpha^i - A_\alpha^i - A_m^i \sigma_\alpha^m) \partial_i,\end{aligned}\quad (5.41)$$

on the composite manifold Y , where D_γ is the covariant differential (5.29) relative to the composite connection (5.40). We call \tilde{D} the *vertical covariant differential*.

5.3.1 Principal Connections

The above general approach to *connections as jet fields* is suitable to formulate the classical concept of *principal connections*. In this section, a structure group G of a principal bundle is assumed to be a real finite-dimensional Lie group (of positive dimension $\dim G > 0$).

A principal connection A on a principal bundle $P \rightarrow Q$ is defined to be a G -equivariant global jet field on P such that

$$j^1 R_g \circ A = A \circ R_g$$

for each canonical map (4.31). We have

$$\begin{aligned}A \circ R_g &= j^1 R_g \circ A, & (g \in G), \\ A &= dq^\alpha \otimes (\partial_\alpha + A_\alpha^m(p) e_m), & (p \in P), \\ A_\alpha^m(qg) &= A_\alpha^m(p) adg^{-1}(e_m).\end{aligned}\quad (5.42)$$

A principal connection A determines splitting $TQ \hookrightarrow T^G P$ of the exact sequence (4.38). We will refer to

$$\mathbf{A} = A - \theta_Q = A_\alpha^m dq^\alpha \otimes e_m \quad (5.43)$$

as a *local connection form*.

Let $J^1(Q, P)$ be the 1-jet space of a principal bundle $P \rightarrow Q$ with a structure Lie group G . The jet prolongation

$$J^1(Q, P) \times J^1(Q \times G) \rightarrow J^1(Q, P)$$

of the canonical action (4.31) brings the jet bundle $J^1(Q, P) \rightarrow Q$ into a general affine bundle modelled on the group bundle

$$J^1(Q \times G) = G \times (T^*Q \otimes \mathfrak{g}_l) \quad (5.44)$$

over Q . However, the jet bundle $J^1(Q, P) \rightarrow Q$ fails to be a principal bundle since the group bundle (5.44) is not a trivial bundle over Q in general. At

the same time, $J^1(Q, P)$ is the G principal bundle $C \times P \rightarrow C$ over the quotient

$$C = J^1(Q, P)/G \quad (5.45)$$

of the jet bundle $J^1(Q, P) \rightarrow P$ by the 1-jet prolongations of the canonical maps (4.31).

Let $J^1(Q, P)$ be the 1-jet space of a principal G -bundle $P \rightarrow Q$. Its quotient (5.45) by the jet prolongation of the canonical action R_G (4.31) is a fibre bundle over Q .

Given a bundle atlas of P and the associated bundle atlas of V_GP , the affine bundle C admits affine bundle coordinates (t, q^i, a_α^q) , and its elements are represented by T_GP -valued 1-forms

$$a = dq^\alpha \otimes (\partial_\alpha + a_\alpha^q e_q) \quad (5.46)$$

on Q . One calls C (5.45) the connection bundle because its sections are naturally identified with principal connections on the principal bundle $P \rightarrow Q$ as follows.

There is the 1-1 correspondence between the principal connections on a principal bundle $P \rightarrow Q$ and the global sections of the quotient bundle

$$C = J^1(Q, P)/G \rightarrow Q.$$

We shall call C the principal connection bundle. It is an affine bundle modelled on the vector bundle

$$\overline{C} = T^*Q \otimes V^G P, \quad (5.47)$$

and there is the canonical vertical splitting

$$VC = C \times \overline{C}.$$

Given a bundle atlas Ψ^P of P , the principal connection bundle C admits the fibre coordinates (q^μ, k_μ^m) so that

$$(k_\mu^m \circ A)(q) = A_\mu^m(q)$$

are coefficients of the local connection one-form (5.43). The 1-jet space $J^1(Q, C)$ of C is with the adapted coordinates

$$(q^\mu, k_\mu^m, k_{\mu\lambda}^m). \quad (5.48)$$

The affine jet bundle $J^1(Q, C) \rightarrow C$ is modelled on the vector bundle

$$T^*Q \otimes (C \times T^*Q \otimes V^G P).$$

There exists the canonical splitting

$$J^1(Q, C) = (J^2 P / G) \otimes (\wedge^2 T^*Q \otimes V^G P) \quad (5.49)$$

over C where

$$C_- = C \times \wedge^2 T^*Q \otimes V^G P$$

and $C_+ \rightarrow C$ is the affine bundle modelled on the vector bundle

$$\overline{C}_+ = \wedge^2 T^*Q \otimes V^G P.$$

In the coordinates (5.48), the splitting (5.49) reads

$$k_{\mu\lambda}^m = \frac{1}{2}(k_{\mu\lambda}^m + k_{\lambda\mu}^m + c_{nl}^m k_\alpha^n k_\mu^l) + \frac{1}{2}(k_{\mu\lambda}^m - k_{\lambda\mu}^m - c_{nl}^m k_\alpha^n k_\mu^l)$$

where c_{mn}^k are structure constants of the Lie algebra \mathfrak{g}_r with respect to its basis $\{I_m\}$.

There are the corresponding canonical projections given by

$$\begin{aligned} \mathcal{S} &= \pi_1 : J^1(Q, C) \rightarrow C_+, & \mathcal{S}_{\lambda\mu}^m &= k_{\mu\lambda}^m + k_{\lambda\mu}^m + c_{nl}^m k_\alpha^n k_\mu^l, \\ \text{and } \mathcal{F} &= \pi_2 : J^1(Q, C) \rightarrow C_-, & \text{with} \\ \mathcal{F} &= \frac{1}{2} \mathcal{F}_{\lambda\mu}^m dq^\alpha \wedge dq^\mu \otimes I_m, & \mathcal{F}_{\lambda\mu}^m &= k_{\mu\lambda}^m - k_{\lambda\mu}^m - c_{nl}^m k_\alpha^n k_\mu^l. \end{aligned}$$

For every principal connection A , we observe that

$$\mathcal{F} \circ j^1 A = F, \quad F = \frac{1}{2} F_{\lambda\mu}^m dq^\alpha \wedge dq^\mu \otimes I_m, \quad F_{\lambda\mu}^m = \partial_\alpha A_\mu^m - \partial_\mu A_\alpha^m - c_{nk}^m A_\alpha^n A_\mu^k,$$

is the strength of A .

Given a symmetric linear connection K^* on the cotangent bundle T^*Q of Q , every principal connection A on a principal bundle P induces the connection

$$S_A : C \rightarrow C_+, \quad S_A \circ A = \mathcal{S} \circ j^1 A,$$

on the principal connection bundle C . In the coordinates (5.48), the connection S_A reads

$$S_{A\mu\lambda}^m = \frac{1}{2} [c_{nl}^m k_\alpha^n k_\mu^l + \partial_\mu A_\alpha^m + \partial_\alpha A_\mu^m - c_{nl}^m (k_\mu^n A_\alpha^l + k_\alpha^n A_\mu^l)] - K^\beta{}_{\mu\lambda} (A_\beta^m - k_\beta^m). \quad (5.50)$$

The P -associated bundle Y admits atlases $\Psi = \{U_\xi, \psi_\xi\}$ associated with atlases $\Psi^P = \{U_\xi, z_\xi\}$ of the principal bundle P as follows:

$$\psi_\xi^{-1}(q \times V) = [z_\xi(q)]_V(V), \quad (q \in U_\xi),$$

where by $[p]_V$ is denoted the restriction of the canonical map $P \times V \rightarrow Y$ to the subset $p \times V$.

Every principal connection A on a principal bundle P induces the associated connection Γ on a P -associated bundle Y such that the following diagram commutes:

$$\begin{array}{ccc} J^1(Q, P) \times V & \longrightarrow & J^1(X, Y) \\ A \times \text{Id}_V \downarrow & & \downarrow \Gamma \\ P \times V & \longrightarrow & Y \end{array}$$

We call it the associated principal connection. With respect to the associated atlases Ψ of Y and Ψ^P of P , this connection is written

$$\Gamma = dq^\alpha \otimes [\partial_\alpha + A_\mu^m(q) I_m^i{}_j y^j \partial_i] \quad (5.51)$$

where $A_\mu^m(q)$ are coefficients of the local connection one-form (5.43) and I_m are generators of the structure group G on the standard fibre V of the bundle Y . The curvature of the connection (5.51) reads

$$R_{\lambda\mu}^i = F_{\lambda\mu}^m I_m^i{}_j y^j.$$

5.4 Definition of a 2-Jet Space

As introduced above, a 2-jet is defined as a second-order equivalence class of functions having the same value and the same first derivatives at some designated point of the domain manifold. Recall that in mechanical settings, the 2-jets are local coordinate maps

$$j_t^2 s : t \mapsto (t, x^i, \dot{x}^i, \ddot{x}^i).$$

In general, if we recursively apply the jet functor $\mathfrak{J} : \text{Bun} \rightarrow \text{Jet}$ to the jet bundles, we come to the higher order jet spaces (see [Kolar *et al.* (1993); Sardanashvily (1993); Sardanashvily (1995); Giachetta *et. al.* (1997); Mangiarotti and Sardanashvily (2000a); Sardanashvily (2002a)]).

In particular, taking the 1-jet space of the 1-jet bundle $J^1(X, Y) \rightarrow X$, we get the *repeated jet space* $J^1(X, J^1(X, Y))$, which admits the adapted coordinates

$$(x^\alpha, y^i, y_\alpha^i, \hat{y}_\mu^i, y_{\mu\alpha}^i)$$

with transition functions

$$\hat{y}'^i_\alpha = \frac{\partial x^\alpha}{\partial x'^\alpha} d_\alpha y'^i, \quad y'^i_{\mu\alpha} = \frac{\partial x^\alpha}{\partial x'^\mu} d_\alpha y'^i_\alpha, \quad d_\alpha = \partial_\alpha + \hat{y}^j_\alpha \partial_j + y^j_{\nu\alpha} \partial^j_\nu.$$

The 2-jet space $J^2(X, Y)$ of a fibre bundle $Y \rightarrow X$ is coordinated by $(x^\alpha, y^i, y_\alpha^i, y_{\alpha\mu}^i)$, with the local symmetry condition $y_{\alpha\mu}^i = y_{\mu\alpha}^i$. The manifold $J^2(X, Y)$ is defined as the set of equivalence classes $j_x^2 s$ of sections $s^i : X \rightarrow Y$ of the bundle $Y \rightarrow X$, which are identified by their values $s^i(x)$ and the values of their first and second-order partial derivatives at points $x \in X$, respectively,

$$y_\alpha^i(j_x^2 s) = \partial_\alpha s^i(x), \quad y_{\alpha\mu}^i(j_x^2 s) = \partial_\alpha \partial_\mu s^i(x).$$

In other words, the 2-jets $j_x^2 s : x^\alpha \mapsto (x^\alpha, y^i, y_\alpha^i, y_{\alpha\mu}^i)$, which are second-order equivalence classes of sections of the fibre bundle $Y \rightarrow X$, can be identified with their codomain set of adapted coordinates on $J^2(X, Y)$,

$$j_x^2 s \equiv (x^\alpha, y^i, y_\alpha^i, y_{\alpha\mu}^i).$$

Let s be a section of a fibre bundle $Y \rightarrow X$, and let $j^1 s$ be its 1-jet prolongation to a section of the jet bundle $J^1(X, Y) \rightarrow X$. The latter induces the section $j^1 j^1 s$ of the repeated jet bundle $J^1(X, J^1(X, Y)) \rightarrow X$. This section takes its values into the 2-jet space $J^2(X, Y)$. It is called the *2-jet prolongation* of the section s , and is denoted by $j^2 s$.

We have the following affine bundle monomorphisms

$$J^2(X, Y) \hookrightarrow \hat{J}^2(X, Y)(X, Y) \hookrightarrow J^1(X, J^1(X, Y))$$

over $J^1(X, Y)$ and the canonical splitting

$$\hat{J}^2(X, Y)(X, Y) = J^2(X, Y) \oplus (\wedge^2 T^* X \otimes VY), \quad \text{given locally by}$$

$$y_{\alpha\mu}^i = \frac{1}{2}(y_{\alpha\mu}^i + y_{\mu\alpha}^i) + \left(\frac{1}{2}(y_{\alpha\mu}^i - y_{\mu\alpha}^i)\right).$$

In particular, the repeated jet prolongation $j^1 j^1 s$ of a section $s : X \rightarrow Y$ of the fibre bundle $Y \rightarrow X$ is a section of the jet bundle $J^1(X, J^1(X, Y)) \rightarrow$

X . It takes its values into $J^2(X, Y)$ and consists with the 2-jet prolongation j^2s of s :

$$j^1 j^1 s(x) = j^2 s(x) = j_x^2 s.$$

Given a 2-jet space $J^2(X, Y)$ of the fibre bundle $Y \rightarrow X$, we have

(i) the fibred map $r_2 : J^2(Y, TY) \rightarrow TJ^2(X, Y)$, given locally by

$$(\dot{y}_\alpha^i, \dot{y}_{\alpha\mu}^i) \circ r_2 = ((\dot{y}^i)_\alpha - y_\mu^i \dot{x}_\alpha^\mu, (\dot{y}^i)_{\alpha\mu} - y_\mu^i \dot{x}_{\alpha\mu}^\mu - y_{\alpha\mu}^i \dot{x}_\alpha^\mu),$$

where $J^2(Y, TY)$ is the 2-jet space of the tangent bundle TY , and

(ii) the canonical isomorphism $VJ^2(X, Y) = J^2(X, VY)$, where $VJ^2(X, Y)$ is the vertical tangent bundle of the fibre bundle $J^2(X, Y) \rightarrow X$, and $J^2(X, VY)$ is the 2-jet space of the fibre bundle $VY \rightarrow X$.

As a consequence, every vector-field u on a fibre bundle $Y \rightarrow X$ admits the 2-jet lift to the projectable vector-field

$$j^2 u = r_2 \circ j^2 u : J^2(X, Y) \rightarrow TJ^2(X, Y).$$

In particular, if $u = u^\alpha \partial_\alpha + u^i \partial_i$ is a projectable vector-field on Y , its 2-jet lift reads

$$\begin{aligned} j^2 u &= u^\alpha \partial_\alpha + u^i \partial_i + (\partial_\alpha u^i + y_\alpha^j \partial_j u^i - y_\mu^i \partial_\alpha u^\mu) \partial_i^\alpha \\ &\quad + [(\partial_\alpha + y_\alpha^j \partial_j + y_{\beta\alpha}^j \partial_j^\beta)(\partial_\alpha + y_\alpha^k \partial_k) u^i - y_\mu^i \dot{x}_{\alpha\beta}^\mu - y_{\mu\beta}^i \dot{x}_\alpha^\mu] \partial_i^{\alpha\beta}. \end{aligned} \quad (5.52)$$

Generalizations of the contact and complementary maps (5.5–5.6) to the 2-jet space $J^2(X, Y)$ read

$$\begin{aligned} \lambda : J^2(X, Y) &\rightarrow T^*X \otimes TJ^1(X, Y) \quad \text{is locally given by} \\ \lambda &= dx^\alpha \otimes \widehat{\partial}_\alpha = dx^\alpha \otimes (\partial_\alpha + y_\alpha^i \partial_i + y_{\mu\alpha}^i \partial_i^\mu), \quad \text{while} \end{aligned} \quad (5.53)$$

$$\begin{aligned} \theta : J^2(X, Y) &\rightarrow T^*J^1(X, Y) \otimes VJ^1(X, Y) \quad \text{is locally given by} \\ \theta &= (dy^i - y_\alpha^i dx^\alpha) \otimes \partial_i + (dy_\mu^i - y_{\mu\alpha}^i dx^\alpha) \otimes \partial_i^\mu. \end{aligned} \quad (5.54)$$

The contact map (5.53) defines the canonical horizontal splitting of the exact sequence

$$0 \rightarrow VJ^1(X, Y) \hookrightarrow TJ^1(X, Y) \rightarrow J^1(X, Y) \times TX \rightarrow 0.$$

Hence, we get the canonical horizontal splitting of a projectable vector-field $j^1 u \equiv \bar{u}$ on $J^1(X, Y)$ over $J^2(X, Y)$:

$$j^1 u = u_H + u_V = u^\alpha [\partial_\alpha + y_\alpha^i \partial_i + y_{\mu\alpha}^i \partial_i^\mu] + [(u^i - y_\alpha^i u^\alpha) \partial_i + (u_\mu^i - y_{\mu\alpha}^i u^\alpha) \partial_i^\mu].$$

Building on the maps (5.53) and (5.54), one can get the horizontal splittings of the canonical tangent-valued 1-form on $J^1(X, Y)$,

$$\theta_{J^1(X, Y)} = dx^\alpha \otimes \partial_\alpha + dy^i \otimes \partial_i + dy_\mu^i \otimes \partial_i^\mu = \alpha + \theta$$

and the associated exterior differential

$$d = d_{\theta_{J^1(X, Y)}} = d_\alpha + d_\theta = d_H + d_V. \quad (5.55)$$

They are similar to the horizontal splittings (5.15) and (5.16).

A 2-jet field (resp. a 2-connection) $\bar{\Gamma}$ on a fibre bundle $Y \rightarrow X$ is defined to be a 1-jet field (resp. a 1-connection) on the jet bundle $J^1(X, Y) \rightarrow X$, i.e., $\bar{\Gamma}$ is a section (resp. a global section) of the bundle $J^1(X, J^1(X, Y)) \rightarrow J^1(X, Y)$.

In the coordinates $(y_\alpha^i, y_{(\mu)}^i, y_{\alpha\mu}^i)$ of the repeated jet space $J^1(X, J^1(X, Y))$, a 2-jet field $\bar{\Gamma}$ is given by the expression

$$(y_\alpha^i, y_{(\mu)}^i, y_{\alpha\mu}^i) \circ \bar{\Gamma} = (y_\alpha^i, \bar{\Gamma}_{(\mu)}^i, \bar{\Gamma}_{\alpha\mu}^i).$$

Using the contact map (5.53), one can represent it by the tangent-valued horizontal 1-form on the jet bundle $J^1(X, Y) \rightarrow X$,

$$\bar{\Gamma} = dx^\mu \otimes (\partial_\mu + \bar{\Gamma}_{(\mu)}^i \partial_i + \bar{\Gamma}_{\alpha\mu}^i \partial_i^\alpha). \quad (5.56)$$

A 2-jet field $\bar{\Gamma}$ on a fibre bundle $Y \rightarrow X$ is called a *sesquiholonomic* (resp. *holonomic*) 2-jet field if it takes its values into the subbundle $\hat{J}^2(X, Y)$ (resp. $J^2(X, Y)$) of $J^1(X, J^1(X, Y))$. We have the coordinate equality $\bar{\Gamma}_{(\mu)}^i = y_{\mu}^i$ for a sesquiholonomic 2-jet field and the additional equality $\bar{\Gamma}_{\alpha\mu}^i = \bar{\Gamma}_{\mu\alpha}^i$ for a holonomic 2-jet field.

Given a symmetric connection K on the cotangent bundle T^*X , every connection Γ on a fibre bundle $Y \rightarrow X$ induces the connection

$$j\Gamma = dx^\mu \otimes [\partial_\mu + \Gamma_\mu^i \partial_i + (\partial_\alpha \Gamma_\mu^i + \partial_j \Gamma_\mu^i y_\alpha^j - K_{\alpha\mu}^\alpha (y_\alpha^i - \Gamma_\alpha^i)) \partial_i^\alpha]$$

on the jet bundle $J^1(X, Y) \rightarrow X$. Note that the curvature R of a connection Γ on a fibre bundle $Y \rightarrow X$ induces the soldering form $\bar{\sigma}_R$ on $J^1(X, Y) \rightarrow X$,

$$\bar{\sigma}_R = R_{\alpha\mu}^i dx^\mu \otimes \partial_i^\alpha.$$

5.5 Higher-Order Jet Spaces

The notion of 1- and 2-jet spaces is naturally extended to higher-order jet spaces. The k -jet space $J^k(X, Y)$ of a fibre bundle $Y \rightarrow X$ is defined as the disjoint union of the equivalence classes $j_x^k s$ of sections $s^i : X \rightarrow Y$ of the fibre bundle $Y \rightarrow X$, identified by their values and the values of the first k terms of their Taylor-series expansion at points x^i in the domain (base) manifold X . $J^k(X, Y)$ is a smooth manifold with the adapted coordinates $(x^\alpha, y_{\alpha_k \dots \alpha_1}^i)$, where

$$y_{\alpha_k \dots \alpha_1}^i(j_x^k s) = \partial_{\alpha_k} \dots \partial_{\alpha_1} s^i(x), \quad (0 \leq k \leq k).$$

The transformation law of these coordinates reads

$$y_{\alpha + \alpha_k \dots \alpha_1}^i = \frac{\partial x^\mu}{\partial' x^\alpha} d_\mu y_{\alpha_k \dots \alpha_1}^i, \quad (5.57)$$

where $\alpha + \alpha_k \dots \alpha_1 = (\alpha \alpha_k \dots \alpha_1)$ and

$$d_\alpha = \partial_\alpha + \sum_{|\alpha_k \dots \alpha_1| < k} y_{\alpha + \alpha_k \dots \alpha_1}^i \partial_i^{\alpha_k \dots \alpha_1} = \partial_\alpha + y_\alpha^i \partial_i + y_{\alpha\mu}^i \partial_i^\mu + \dots$$

are higher-order total derivatives. These derivatives act on exterior forms on $J^k(X, Y)$ and obey the relations

$$d_\alpha(\phi \wedge \sigma) = d_\alpha(\phi) \wedge \sigma + \phi \wedge d_\alpha(\sigma), \quad d_\alpha(d\phi) = d(d_\alpha(\phi)).$$

For example,

$$d_\alpha(dx^\mu) = 0, \quad d_\alpha(dy_{\alpha_k \dots \alpha_1}^i) = dy_{\alpha + \alpha_k \dots \alpha_1}^i.$$

Let us also mention the following two operations: the horizontal projection h_0 given by the relations

$$h_0(dx^\alpha) = dx^\alpha, \quad h_0(dy_{\alpha_k \dots \alpha_1}^i) = y_{\mu \alpha_k \dots \alpha_1}^i dx^\mu, \quad (5.58)$$

and the horizontal differential

$$d_H(\phi) = dx^\alpha \wedge d_\alpha(\phi), \quad d_H \circ d_H = 0, \quad h_0 \circ d = d_H \circ h_0.$$

In a similar way, one can describe the *infinite-order jet space*, $J^\infty(X, Y)$, which can be coordinated by $(x^\alpha, y^i, \dots, y_{\alpha_1 \dots \alpha_r}^i, \dots)$, where $\alpha_1 \dots \alpha_r$ are collections of numbers modulo rearrangements, but it fails to be a well-behaved manifold in general. At the same time, one can introduce the sheaf of smooth functions on $J^\infty(X, Y)$ and define the differential calculus

on $J^\infty(X, Y)$, with suitable notation for vector-fields, derivatives and differential forms just as like as in the finite order case (see [Sardanashvily (1993); Sardanashvily (1995); Giachetta *et. al.* (1997); Mangiarotti and Sardanashvily (2000a); Sardanashvily (2002a)]).

A vector-field u_k on the k -jet space $J^k(X, Y)$ is called *projectable vector-field* if for any $l < k$ there exists a vector-field u_l on $J^l(X, Y) \rightarrow X$ such that

$$u_l \circ \pi_l^k = T\pi_l^k \circ u_k.$$

The tangent map $T\pi_l^k$ sends projectable vector-fields on $J^k(X, Y)$ onto the projectable vector-fields on $J^l(X, Y)$.

Now consider projectable vector-fields u_k which are extension to the higher-order jet spaces of infinitesimal transformations of the fibre bundle $Y \rightarrow X$. The linear space of projectable vector-fields on $J^\infty(X, Y)$ is defined as the limit of the inverse system of projectable vector-fields on k -jet spaces. As a consequence, every projectable vector-field on the bundle $Y \rightarrow X$,

$$u = u^\alpha \partial_\alpha + u^i \partial_i,$$

induces a projectable vector-field \bar{u}^∞ on $J^\infty(X, Y)$. We have its canonical decomposition

$$\begin{aligned} \bar{u}^\infty &= u_H^\infty + u_V^\infty, \\ u_H^\infty &= u^\alpha \hat{\partial}_\alpha^\infty = u^\alpha (\partial_\alpha + y_\alpha^i \partial_i + \dots), \\ u_V^\infty &= \sum_{k=0}^{\infty} \hat{\partial}_{\alpha_k}^k \dots \hat{\partial}_{\alpha_1}^1 u_V^i \partial_i^{\alpha_1 \dots \alpha_k}, \end{aligned} \quad (5.59)$$

where u_V is the vertical part of the splitting (5.14) of $\pi_0^* u$. In particular, u_H^∞ is the canonical lift of the vector-field $\tau = u^\alpha \partial_\alpha$ on X onto $J^\infty(X, Y)$.

By the same limiting process we can introduce the notions of inner product of exterior forms and projectable vector-fields, the Lie bracket of projectable vector-fields and the Lie derivative of exterior forms by projectable vector-fields on $J^\infty(X, Y)$. All the usual identities are satisfied.

In particular, the notion of contact forms is extended to the forms

$$\hat{d}y_{\alpha_1 \dots \alpha_r}^i = dy_{\alpha_1 \dots \alpha_r}^i - y_{\alpha_1 \dots \alpha_r \nu}^i dx^\nu.$$

Let $\Omega^{r,k}$ denote the space of exterior forms on $J^\infty(X, Y)$ which are of the order r in the horizontal forms dx^ν and of the order k in the contact forms.

Then, the space Ω^n of exterior n -forms on $J^\infty(X, Y)$ admits the unique decomposition

$$\Omega^n = \Omega^{n,0} \oplus \Omega^{n-1,1} \oplus \dots \oplus \Omega^{0,n}. \quad (5.60)$$

An exterior form is called a k -contact form if it belongs to the space $\Omega^{r,k}$. In particular, we have the k -contact projection $h_k : \Omega^n \rightarrow \Omega^{n-k,k}$. For example, the horizontal projection h_0 performs the replacement $dy_{\alpha_1 \dots \alpha_k}^i \rightarrow y_{\alpha_1 \dots \alpha_k \nu}^i dx^\nu$.

The exterior differential d on exterior forms on $J^\infty(X, Y)$ is decomposed into the sum

$$d = d_H + d_V \quad (5.61)$$

of the total differential operator

$$d_H \phi = \widehat{\partial}_\mu^\infty \phi \dots dx^\mu \wedge \dots$$

and the vertical differential operator

$$d_V \phi = \frac{\partial \phi \dots}{\partial y_{\alpha_1 \dots \alpha_r}^i} \widehat{d}y_{\alpha_1 \dots \alpha_r}^i \wedge \dots$$

These differentials satisfy the cohomology properties

$$d_H d_H = 0, \quad d_V d_V = 0, \quad d_V d_H + d_H d_V = 0.$$

Note that if σ is an exterior form on the k -jet space $J^k(X, Y)$, the decomposition (5.61) is reduced to

$$\begin{aligned} \pi_r^{r+1*} d\sigma &= d_H \sigma + d_V \sigma, & \text{which implies} \\ h_0(d\sigma) &= d_H h_0(\sigma). \end{aligned}$$

5.6 Application: Jets and Non-Autonomous Dynamics

As complex nonlinear mechanics is the most exact basis of all complex nonlinear dynamical systems considered in this book, we give here the first glimpse of mechanics on jet spaces.

Recall that in ordinary (autonomous) mechanics we have a *configuration manifold* M and the corresponding *velocity phase-space manifold* is

its tangent bundle TM . However, in modern geometrical settings of non-autonomous (see [Giachetta *et. al.* (1997); Mangiarotti *et. al.* (1999); Mangiarotti and Sardanashvily (2000a); Saunders (1989); Sardanashvily (1993); Sardanashvily (1995); Sardanashvily (2002a)]), the configuration manifold of time-dependent mechanics is a fibre bundle $Q \rightarrow \mathbb{R}$, called the *configuration bundle*, coordinated by (t, q^i) , where $t \in \mathbb{R}$ is a Cartesian coordinate on the time axis \mathbb{R} with the transition functions $t' = t + \text{const}$. The corresponding velocity phase-space is the 1-jet space $J^1(\mathbb{R}, Q)$, which admits the adapted coordinates (t, q^i, \dot{q}^i) . It was proved in [Giachetta (1992); León *et. al.* (1996); Mangiarotti and Sardanashvily (1998)] that every *dynamical equation* ξ defines a *connection* on the affine jet bundle $J^1(\mathbb{R}, Q) \rightarrow Q$, and *vice versa*.

Due to the canonical imbedding $J^1(\mathbb{R}, Q) \rightarrow TQ$, every dynamical connection induces a nonlinear connection on the tangent bundle $TQ \rightarrow Q$, and *vice versa*. As a consequence, every dynamical equation on Q induces an equivalent geodesic equation on the tangent bundle $TQ \rightarrow Q$ in accordance with the following proposition. Given a configuration bundle $Q \rightarrow \mathbb{R}$, coordinated by (t, q^i) , and its 2-jet space $J^2(\mathbb{R}, Q)$, coordinated by $(t, q^i, \dot{q}^i, \ddot{q}^i)$, any dynamical equation ξ on the configuration bundle $Q \rightarrow \mathbb{R}$,

$$q_{tt}^i = \xi^i(t, q^i, \dot{q}^i) \quad (5.62)$$

is equivalent to the *geodesic equation* with respect to a connection \tilde{K} on the tangent bundle $TQ \rightarrow Q$,

$$\dot{t} = 1, \quad \ddot{t} = 0, \quad \ddot{q}^i = \tilde{K}_0^i + \tilde{K}_j^i \dot{q}^j,$$

which fulfills the conditions

$$\tilde{K}_\alpha^0 = 0, \quad \xi^i = \tilde{K}_0^i + q_t^j \tilde{K}_j^i \big|_{t=1, \dot{q}^i = q_t^i}. \quad (5.63)$$

Recall that the 1-jet space $J^1(\mathbb{R}, Q)$ is defined as the set of equivalence classes $j_t^1 c$ of sections $c^i : \mathbb{R} \rightarrow Q$ of the fibre bundle $Q \rightarrow \mathbb{R}$, which are identified by their values $c^i(t)$ and the values of their partial derivatives $\partial_t c^i = \partial_t c^i(t)$ at time points $t \in \mathbb{R}$. Also recall that there is the canonical imbedding

$$\lambda : J^1(\mathbb{R}, Q) \hookrightarrow TQ, \quad \text{locally given by} \quad \lambda = d_t = \partial_t + q_t^i \partial_i, \quad (5.64)$$

where d_t denotes the total time derivative. From now on, we will identify $J^1(\mathbb{R}, Q)$ with its image in the tangent bundle TQ . This is an affine bundle modelled over the vertical tangent bundle VQ of $Q \rightarrow \mathbb{R}$.

As a consequence of (5.64), every connection Γ on a fibre bundle $Q \rightarrow \mathbb{R}$,

$$\Gamma : Q \rightarrow J^1(\mathbb{R}, Q), \quad \text{locally given by} \quad \Gamma = dt \otimes (\partial_t + \Gamma^i \partial_i), \quad (5.65)$$

is identified with the nowhere vanishing vector-field on Q [Mangiarotti and Sardanashevily (1998); Mangiarotti *et. al.* (1999)],

$$\Gamma : Q \rightarrow J^1(\mathbb{R}, Q) \subset TQ, \quad \text{locally given by} \quad \Gamma = \partial_t + \Gamma^i \partial_i. \quad (5.66)$$

This is the horizontal lift of the standard vector-field ∂_t on \mathbb{R} by means of the connection (5.65). Conversely, any vector-field Γ on Q such that $dt \rfloor \Gamma = 1$ defines a connection on $Q \rightarrow \mathbb{R}$. Therefore, the covariant differential associated with a connection Γ on $Q \rightarrow \mathbb{R}$ reads

$$D^G : J^1(\mathbb{R}, Q) \rightarrow VQ, \quad \text{locally given by} \quad q^i \circ D_G = q_t^i - \Gamma^i.$$

Let $J^1(\mathbb{R}, J^1(\mathbb{R}, Q))$ denote the (repeated) 1-jet space of the jet bundle $J^1(\mathbb{R}, Q) \rightarrow \mathbb{R}$, coordinated by $(t, q^i, q_t^i, q_{(t)}^i, q_{tt}^i)$. The corresponding 2-jet space $J^2(\mathbb{R}, Q)$ of the fibre bundle $Q \rightarrow \mathbb{R}$ is the holonomic subbundle $q_t^i = q_{(t)}^i$ of $J^1(\mathbb{R}, J^1(\mathbb{R}, Q))$, coordinated by $(t, q^i, q_t^i, q_{tt}^i)$. There are the imbeddings

$$J^2(\mathbb{R}, Q) \xrightarrow{\tilde{\lambda}} TJ^1(\mathbb{R}, Q) \xrightarrow{T\lambda} TTQ, \quad \text{with} \\ \tilde{\lambda} : (t, q^i, q_t^i, q_{tt}^i) \mapsto (t, q^i, q_t^i, \dot{t} = 1, \dot{q}^i = q_t^i, \dot{q}_t^i = q_{tt}^i). \quad (5.67)$$

$$T\lambda \circ \tilde{\lambda} : (t, q^i, q_t^i, q_{tt}^i) \mapsto (t, q^i, \dot{t} = 1, \dot{q}^i = q_t^i, \ddot{t} = 0, \ddot{q}^i = q_{tt}^i), \quad (5.68)$$

where $(t, q^i, \dot{q}^i, \ddot{q}^i)$ are holonomic coordinates on the second tangent bundle TTQ . This global geometrical structure of time-dependent mechanics is depicted in Figure 5.3.

Therefore, a dynamical equation ξ on a configuration bundle $Q \rightarrow \mathbb{R}$, given in local coordinates by (5.62), is defined as the *geodesic equation* $\text{Ker } D^\xi \subset J^2(\mathbb{R}, Q)$ for a holonomic connection ξ on the jet bundle $J^1(\mathbb{R}, Q) \rightarrow \mathbb{R}$. Due to the map (5.67), a holonomic connection ξ is represented by the horizontal vector-field on $J^1(\mathbb{R}, Q)$,

$$\xi = \partial_t + q_t^i \partial_i + \xi^i(q^\mu, q_t^\mu) \partial_t^i. \quad (5.69)$$

A dynamical equation ξ is said to be *conservative* if there exists a trivialization $Q \cong \mathbb{R} \times M$ such that the vector-field ξ (5.69) on $J^1(\mathbb{R}, Q) \cong \mathbb{R} \times TM$ is projectable onto TM . Then this projection

$$\Xi_\xi = \dot{q}^i \partial_i + \xi^i(q^j, \dot{q}^j) \dot{\partial}_i$$

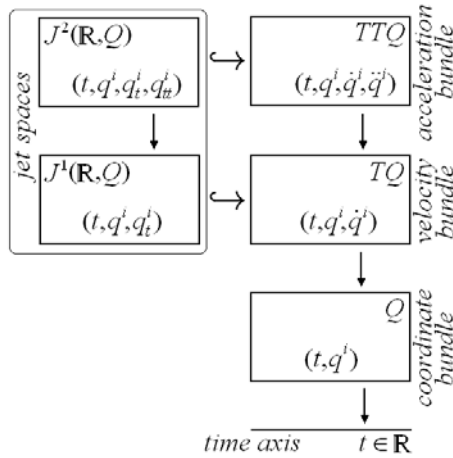


Fig. 5.3 Hierarchical geometrical structure of time-dependent mechanics. Note that (for simplicity) intermediate jet spaces, $J^1(\mathbb{R}, J^1(\mathbb{R}, Q))$ and $TJ^1(\mathbb{R}, Q)$, are not shown.

is a second-order dynamical equation on a typical fibre M of $Q \rightarrow \mathbb{R}$,

$$\ddot{q}^i = \Xi_{\xi}^i. \quad (5.70)$$

Conversely, every second-order dynamical equation Ξ (5.70) on a manifold M can be seen as a conservative dynamical equation

$$\xi_{\Xi} = \partial_t + \dot{q}^i \partial_i + u^i \partial_i$$

on the trivial fibre bundle $\mathbb{R} \times M \rightarrow \mathbb{R}$.

Now we can explore the fundamental relationship between the *holonomic connections* ξ (5.69) on the 1-jet bundle $J^1(\mathbb{R}, Q) \rightarrow \mathbb{R}$ and the *dynamical connections* γ on the *affine 1-jet bundle* $J^1(\mathbb{R}, Q) \rightarrow Q$, given by

$$\gamma = dq^{\alpha} \otimes (\partial_{\alpha} + \gamma_{\alpha}^i \partial_i), \quad (q^{\alpha} \equiv (t, q^i), \partial_{\alpha} \equiv (\partial_t, \partial_i)). \quad (5.71)$$

Any dynamical connection γ (5.71) defines the holonomic connection ξ_{γ} on $J^1(\mathbb{R}, Q) \rightarrow \mathbb{R}$ [Mangiarotti and Sardanashvily (1998)]

$$\xi_{\gamma} = \partial_t + q_t^i \partial_i + (\gamma_0^i + q_t^j \gamma_j^i) \partial_i.$$

Conversely, any holonomic connection ξ (5.69) on $J^1(\mathbb{R}, Q) \rightarrow \mathbb{R}$ defines the dynamical connection

$$\gamma_{\xi} = dt \otimes [\partial_t + (\xi^i - \frac{1}{2} q_t^j \partial_j^t \xi^i) \partial_i] + dq^j \otimes [\partial_j + \frac{1}{2} \partial_j^t \xi^i \partial_i^t]. \quad (5.72)$$

It follows that every dynamical connection γ (5.71) induces the dynamical equation (5.62) on the configuration bundle $Q \rightarrow \mathbb{R}$, rewritten here as

$$q_{tt}^i = \gamma_0^i + q_t^j \gamma_j^i. \quad (5.73)$$

Different dynamical connections may lead to the same dynamical equation (5.73). The dynamical connection γ_ξ (5.72), associated with a dynamical equation, possesses the property

$$\gamma_i^k = \partial_i^t \gamma_0^k + q_t^j \partial_i^t \gamma_j^k,$$

which implies the relation $\partial_j^t \gamma_i^k = \partial_i^t \gamma_j^k$. Such a dynamical connection is called symmetric. Let γ be a dynamical connection (5.71) and ξ_γ the corresponding dynamical equation (5.6). Then the connection (5.72), associated with ξ_γ , takes the form

$$\gamma_{\xi_\gamma}^k{}_i = \frac{1}{2}(\gamma_i^k + \partial_i^t \gamma_0^k + q_t^j \partial_i^t \gamma_j^k), \quad \gamma_{\xi_\gamma}^k{}_0 = \xi^k - q_t^i \gamma_{\xi_\gamma}^k{}_i.$$

Note that $\gamma = \gamma_{\xi_\gamma}$ iff γ is symmetric.

To explore the relation between the connections γ (5.71) on the affine jet bundle $J^1(\mathbb{R}, Q) \rightarrow Q$ and the connections

$$K = dq^\alpha \otimes (\partial_\alpha + K_\alpha^\beta \dot{\partial}_\beta) \quad (5.74)$$

on the tangent bundle $TQ \rightarrow Q$, consider the diagram

$$\begin{array}{ccc} J^1(\mathbb{R}, J^1(\mathbb{R}, Q)) & \xrightarrow{j^1\lambda} & J^1(Q, TQ) \\ \gamma \downarrow & & \downarrow K \\ J^1(\mathbb{R}, Q) & \xrightarrow{\lambda} & TQ \end{array} \quad (5.75)$$

where $J^1(Q, TQ)$ is the 1-jet space of the tangent bundle $TQ \rightarrow Q$, coordinated by $(q^\alpha, \dot{q}^\alpha, \dot{q}_\mu^\alpha)$. The jet prolongation $j^1\lambda$ of the canonical imbedding λ (5.64) over Q reads

$$j^1\lambda : (t, q^i, q_t^i, q_{\mu t}^i) \mapsto (t, q^i, \dot{t} = 1, \dot{q}^i = q_t^i, \dot{t}_\mu = 0, \dot{q}_\mu^i = q_{\mu t}^i).$$

We have

$$\begin{aligned} j^1\lambda \circ \gamma : (t, q^i, q_t^i) &\mapsto (t, q^i, \dot{t} = 1, \dot{q}^i = q_t^i, \dot{t}_\mu = 0, \dot{q}_\mu^i = \gamma_\mu^i), \\ K \circ \lambda : (t, q^i, q_t^i) &\mapsto (t, q^i, \dot{t} = 1, \dot{q}^i = q_0^i, \dot{t}_\mu = K_\mu^0, \dot{q}_\mu^i = K_\mu^i). \end{aligned}$$

It follows that the diagram (5.75) can be commutative only if the components K_μ^0 of the connection K on $TQ \rightarrow Q$ vanish. Since the transition functions $t \rightarrow t'$ are independent of q^i , a connection

$$\tilde{K} = dq^\alpha \otimes (\partial_\alpha + K_\alpha^i \partial_i) \quad (5.76)$$

with the components $K_\mu^0 = 0$ can exist on the tangent bundle $TQ \rightarrow Q$. It obeys the transformation law

$$K_\alpha^i = (\partial_j x'^i K_\mu^j + \partial_\mu x'^i) \frac{\partial q^\mu}{\partial x'^\alpha}. \quad (5.77)$$

Now the diagram (5.75) becomes commutative if the connections γ and \tilde{K} fulfill the relation

$$\gamma_\mu^i = K_\mu^i(t, q^i, \dot{t} = 1, \dot{q}^i = q_t^i),$$

which holds globally since the substitution of $\dot{q}^i = q_t^i$ into (5.77) restates the coordinate transformation law of γ .

Every dynamical equation (5.62) on the configuration bundle $Q \rightarrow \mathbb{R}$ can be written in the form

$$q_{tt}^i = K_0^i \circ \lambda + q_t^j K_j^i \circ \lambda, \quad (5.78)$$

where \tilde{K} is a connection (5.76) on the tangent bundle $TQ \rightarrow Q$. Conversely, each connection \tilde{K} (5.76) on $TQ \rightarrow Q$ defines the dynamical equation (5.78) on $Q \rightarrow \mathbb{R}$.

Consider the geodesic equation (5.6) on TQ with respect to the connection \tilde{K} . Its solution is a geodesic curve $c(t)$ which also satisfies the dynamical equation (5.62), and *vice versa*.

From the physical viewpoint, a reference frame in mechanics on a configuration bundle $Q \rightarrow \mathbb{R}$ sets a tangent vector at each point of Q which characterizes the velocity of an 'observer' at this point. Then any connection Γ on $Q \rightarrow \mathbb{R}$ is said to be such a reference frame [Echeverría *et. al.* (1995); Mangiarotti and Sardanashvily (1998); Massa and Pagani (1994); Sardanashvily (1998)].

Each connection Γ on a fibre bundle $Q \rightarrow \mathbb{R}$ defines an atlas of local constant trivializations of $Q \rightarrow \mathbb{R}$ whose transition functions are independent of t , and *vice versa*. One finds $\Gamma = \partial_t$ with respect to this atlas. In particular, there is 1–1 correspondence between the complete connections Γ (5.66) on $Q \rightarrow \mathbb{R}$ and the trivializations of this bundle.

Given a reference frame Γ , any connection K (5.74) on the tangent bundle $TQ \rightarrow Q$ defines the dynamical equation

$$\xi^i = (K_\alpha^i - \Gamma^i K_\alpha^0) \dot{q}^\alpha \big|_{i=1, \dot{q}^i = q_t^i}. \quad (5.79)$$

Given a connection Γ on the fibre bundle $Q \rightarrow \mathbb{R}$ and a connection K on the tangent bundle $TQ \rightarrow Q$, there is the connection \tilde{K} on $TQ \rightarrow Q$ with the components

$$\tilde{K}_\alpha^0 = 0, \quad \tilde{K}_\alpha^i = K_\alpha^i - \Gamma^i K_\alpha^0.$$

5.6.1 Geodesics

In this subsection we continue our study of non-autonomous, *time-dependent mechanics* on a *configuration bundle* $Q \rightarrow \mathbb{R}$, that we started in subsection 5.6 above. Recall that \mathbb{R} is the time axis, while the corresponding *velocity phase-space manifold* is the 1-jet space $J^1(\mathbb{R}, Q)$ of sections $s: \mathbb{R} \rightarrow Q$ of the bundle $Q \rightarrow \mathbb{R}$. Also, recall that second-order dynamical equation (dynamical equation, for short) on a fibre bundle $Q \rightarrow \mathbb{R}$ is defined as a first-order dynamical equation on the jet bundle $J^1(\mathbb{R}, Q) \rightarrow \mathbb{R}$, given by a holonomic connection ξ on $J^1(\mathbb{R}, Q) \rightarrow \mathbb{R}$ which takes its values in the 2-jet space $J^2(\mathbb{R}, Q) \subset J^1(\mathbb{R}, J^1(\mathbb{R}, Q))$ (see [León *et. al.* (1996); Mangiarotti and Sardanashvily (1998); Massa and Pagani (1994); Mangiarotti and Sardanashvily (1999)]). The global geometrical structure of time-dependent mechanics is depicted in Figure 5.3 above.

Since a configuration bundle $Q \rightarrow \mathbb{R}$ is trivial, the existent formulations of mechanics often imply its preliminary splitting $Q = \mathbb{R} \times M$ [Cariñena and F.Núñez (1993); Echeverría *et. al.* (1991); León *et. al.* (1996); Morandi *et. al.* (1990)]. This is not the case of mechanical systems subject to time-dependent transformations, including inertial frame transformations. Recall that different trivializations of $Q \rightarrow \mathbb{R}$ differ from each other in projections $Q \rightarrow M$. Since a configuration bundle $Q \rightarrow \mathbb{R}$ has no canonical trivialization in general, mechanics on $Q \rightarrow \mathbb{R}$ is not a repetition of mechanics on $\mathbb{R} \times M$, but implies additionally a connection on $Q \rightarrow \mathbb{R}$ which is a reference frame [Mangiarotti and Sardanashvily (1998); Sardanashvily (1998)]. Considered independently on a trivialization of $Q \rightarrow \mathbb{R}$, mechanical equations make the geometrical sense of geodesic equations.

We now examine quadratic dynamical equations in details. In this case, the corresponding dynamical connection γ on $J^1(\mathbb{R}, Q) \rightarrow Q$ is

affine, while the connection \tilde{K} (5.63) on $TQ \rightarrow Q$ is *linear*. Then the equation for Jacobi vector-fields along the geodesics of the connection \tilde{K} can be considered. This equation coincides with the existent equation for Jacobi fields of a Lagrangian system [Dittrich and Reuter (1994); Mangiarotti and Sardanashvily (1998)] in the case of non-degenerate quadratic Lagrangians, when they can be compared. We will consider more general case of quadratic Newtonian systems characterized by a pair (ξ, μ) of a quadratic dynamical equation ξ and a Riemannian inertia tensor μ which satisfy a certain compatibility condition. Given a reference frame, a Riemannian inertia tensor μ is extended to a Riemannian metric on the configuration space Q . Then conjugate points of solutions of the dynamical equation ξ can be examined in accordance with the well-known geometrical criteria [Mangiarotti and Sardanashvily (1998); Sardanashvily (1998)].

5.6.2 Quadratic Dynamical Equations

From the physical viewpoint, the most interesting dynamical equations are the quadratic ones, i.e.,

$$\xi^i = a_{jk}^i(q^\mu) q_t^j q_t^k + b_j^i(q^\mu) q_t^j + f^i(q^\mu). \quad (5.80)$$

This property is coordinate-independent due to the affine transformation law of coordinates q_t^i . Then, it is clear that the corresponding dynamical connection γ_ξ (5.72) is affine:

$$\gamma = dq^\alpha \otimes [\partial_\alpha + (\gamma_{\lambda 0}^i(q^\nu) + \gamma_{\lambda j}^i(q^\nu) q_t^j) \partial_i^t],$$

and *vice versa*. This connection is symmetric iff $\gamma_{\lambda\mu}^i = \gamma_{\mu\lambda}^i$.

There is 1-1 correspondence between the affine connections γ on the affine jet bundle $J^1(\mathbb{R}, Q) \rightarrow Q$ and the linear connections \tilde{K} (5.76) on the tangent bundle $TQ \rightarrow Q$.

This correspondence is given by the relation

$$\gamma_\mu^i = \gamma_{\mu 0}^i + \gamma_{\mu j}^i q_t^j, \quad \gamma_{\mu\lambda}^i = K_\mu^i{}_\alpha.$$

In particular, if an affine dynamical connection γ is symmetric, so is the corresponding linear connection \tilde{K} .

Any quadratic dynamical equation

$$q_{tt}^i = a_{jk}^i(q^\mu) q_t^j q_t^k + b_j^i(q^\mu) q_t^j + f^i(q^\mu) \quad (5.81)$$

is equivalent to the geodesic equation

$$\begin{aligned}\dot{t} &= 1, & \ddot{t} &= 0, \\ \ddot{q}^i &= a_{jk}^i(q^\mu) \dot{q}^j \dot{q}^k + b_j^i(q^\mu) \dot{q}^j \dot{t} + f^i(q^\mu) \dot{t} \dot{t}.\end{aligned}\quad (5.82)$$

for the symmetric linear connection

$$\tilde{K} = dq^\alpha \otimes (\partial_\alpha + K_{\alpha\nu}^\mu(t, q^i) \dot{q}^\nu \partial_\mu)$$

on $TQ \rightarrow Q$, given by the components

$$K_{\alpha\nu}^0 = 0, \quad K_0^i{}^0 = f^i, \quad K_0^i{}^j = K_j^i{}^0 = \frac{1}{2}b_j^i, \quad K_j^i{}^k = a_{jk}^i. \quad (5.83)$$

Conversely, any linear connection K on the tangent bundle $TQ \rightarrow Q$ defines the quadratic dynamical equation

$$q_{tt}^i = K_j^i{}^k q_t^j q_t^k + (K_0^i{}^j + K_j^i{}^0) q_t^j + K_0^i{}^0,$$

written with respect to a given reference frame $(t, q^i) \equiv q^\mu$.

However, the geodesic equation (5.82) is not unique for the dynamical equation (5.81). Any quadratic dynamical equation (5.80), being equivalent to the geodesic equation with respect to the linear connection \tilde{K} (5.83), is also equivalent to the geodesic equation with respect to an affine connection K' on $TQ \rightarrow Q$ which differs from \tilde{K} (5.83) in a soldering form σ on $TQ \rightarrow Q$ with the local components

$$\sigma_\alpha^0 = 0, \quad \sigma_k^i = h_k^i + (s-1)h_k^i \dot{q}^0, \quad \sigma_0^i = -sh_k^i \dot{q}^k - h_0^i \dot{q}^0 + h_0^i,$$

where s and h_α^i are local functions on Q .

5.6.3 Equation of Free-Motion

We say that the dynamical equation (5.62), that is: $q_{tt}^i = \xi^i(t, q^i, \dot{q}_t^i)$, is a *free motion equation* iff there exists a reference frame (t, \bar{q}^i) on the configuration bundle $Q \rightarrow \mathbb{R}$ such that this equation reads

$$\bar{q}_{tt}^i = 0. \quad (5.84)$$

With respect to arbitrary bundle coordinates (t, q^i) , a free motion equation takes the form

$$q_{tt}^i = d_t \Gamma^i + \partial_j \Gamma^i (q_t^j - \Gamma^j) - \frac{\partial q^i}{\partial \bar{q}^\mu} \frac{\partial \bar{q}^\mu}{\partial q^j \partial q^k} (q_t^j - \Gamma^j) (q_t^k - \Gamma^k), \quad (5.85)$$

where $\Gamma^i = \partial_t q^i(t, \bar{q}^j)$ is the connection associated with the initial frame (t, \bar{q}^i) . One can think of the r.h.s. of the equation (5.85) as being the general coordinate expression of an inertial force in mechanics. The corresponding dynamical connection γ on the affine jet bundle $J^1(\mathbb{R}, Q) \rightarrow Q$ reads

$$\gamma_k^i = \partial_k \Gamma^i - \frac{\partial q^i}{\partial \bar{q}^\mu} \frac{\partial \bar{q}^\mu}{\partial q^j \partial q^k} (q_t^j - \Gamma^j), \quad \gamma_0^i = \partial_t \Gamma^i + \partial_j \Gamma^i q_t^j - \gamma_k^i \Gamma^k. \quad (5.86)$$

This affine dynamical connection defines a linear connection K on the tangent bundle $TQ \rightarrow Q$ whose curvature is necessarily zero. Thus, we come to the following criterion for a dynamical equation to be a free motion equation: if ξ is a free motion equation, it is quadratic and the corresponding linear symmetric connection (5.83) on the tangent bundle $TQ \rightarrow Q$ is flat. A free motion equation on a configuration bundle $Q \rightarrow \mathbb{R}$ exists iff the typical fibre M of Q admits a curvature-free linear symmetric connection.

5.6.4 Quadratic Lagrangian and Newtonian Systems

Recall that a Lagrangian of an n D mechanical system on $Q \rightarrow \mathbb{R}$ is defined as a function on the velocity phase-space $J^1(\mathbb{R}, Q)$. In particular, let us consider a *non-degenerate quadratic Lagrangian*

$$L = \frac{1}{2} \mu_{ij} (q^\mu) q_t^i q_t^j + k_i (q^\mu) q_t^i + f(q^\mu), \quad (5.87)$$

where μ_{ij} ($i, j = 1, \dots, n$) is a Riemannian fibre metric tensor in the vertical tangent bundle VQ , called the *inertial metric tensor*. As for quadratic dynamical equations, this property is coordinate-independent, namely one can show that any quadratic polynomial on $J^1(\mathbb{R}, Q) \subset TQ$ is extended to a bilinear form on TQ , so that the Lagrangian L (5.87) can be written as

$$L = \frac{1}{2} \gamma_{\alpha\mu} q_t^\alpha q_t^\mu, \quad (\text{with } q_t^0 = 1),$$

where γ is the (degenerate) fibre metric in the tangent bundle TQ , given by

$$\gamma_{00} = 2f, \quad \gamma_{0i} = k_i, \quad \gamma_{ij} = \mu_{ij}. \quad (5.88)$$

The associated Lagrangian equation takes the form

$$q_{tt}^i = (\mu^{-1})^{ik} \Gamma_{\lambda k \nu} q_t^\alpha q_t^\nu, \quad (5.89)$$

$$\text{where } \Gamma_{\lambda \mu \nu} = -\frac{1}{2} (\partial_\alpha \gamma_{\mu \nu} + \partial_\nu \gamma_{\mu \lambda} - \partial_\mu \gamma_{\lambda \nu})$$

are the *Christoffel symbols of the first-kind* of the metric $\gamma_{\alpha\mu}$ given in components by (5.88). The corresponding geodesic equation (5.82) on TQ reads

$$\ddot{q}^i = (\mu^{-1})^{ik} \Gamma_{\lambda k \nu} \dot{q}^\lambda \dot{q}^\nu, \quad \dot{t} = 1, \quad \ddot{t} = 0, \quad (5.90)$$

such that \tilde{K} (5.63) is a linear connection with the following components

$$\tilde{K}_{\alpha\nu}^0 = 0, \quad \tilde{K}_{\alpha\nu}^i = (\mu^{-1})^{ik} \Gamma_{\lambda k \nu}.$$

We have the relation

$$\dot{q}^\alpha (\partial_\alpha \mu_{ij} + K_{\alpha\nu}^i \dot{q}^\nu) = 0. \quad (5.91)$$

One can show that an arbitrary Lagrangian system on a configuration bundle $Q \rightarrow \mathbb{R}$ is a particular Newtonian system on $Q \rightarrow \mathbb{R}$. The latter is defined as a pair (ξ, μ) of a dynamical equation ξ and a (degenerate) fibre metric μ in the fibre bundle $V_Q J^1(\mathbb{R}, Q) \rightarrow J^1(\mathbb{R}, Q)$ which satisfy the symmetry condition $\partial_k^t \mu_{ij} = \partial_j^t \mu_{ik}$ and the compatibility condition [Mangiarotti and Sardanashvily (1998); Mangiarotti *et. al.* (1999)]

$$\xi \rfloor d\mu_{ij} + \mu_{ik} \gamma_j^k + \mu_{jk} \gamma_i^k = 0, \quad (5.92)$$

where γ_ξ is the dynamical connection (5.72), i.e.,

$$\gamma_\xi = dt \otimes [\partial_t + (\xi^i - \frac{1}{2} q_t^j \partial_j^t \xi^i) \partial_i^t] + dq^j \otimes [\partial_j + \frac{1}{2} \partial_j^t \xi^i \partial_i^t].$$

We restrict our consideration to non-degenerate quadratic Newtonian systems when ξ is a quadratic dynamical equation (5.80) and μ is a Riemannian fibre metric in VQ , i.e., μ is independent of q_t^i and the symmetry condition becomes trivial. In this case, the dynamical equation (5.81) is equivalent to the geodesic equation (5.82) with respect a symmetric linear connection \tilde{K} (5.83), while the compatibility condition (5.92) takes the form (5.91).

Given a symmetric linear connection \tilde{K} (5.83) on the tangent bundle $TQ \rightarrow Q$, one can consider the equation for Jacobi vector-fields along geodesics of this connection, i.e., along solutions of the dynamical equation (5.81). If Q admits a Riemannian metric, the *conjugate points* of these geodesic can be investigated.

5.6.5 Jacobi Fields

Let us consider the quadratic dynamical equation (5.81) and the equivalent geodesic equation (5.82) with respect to the symmetric linear connection \tilde{K} (5.83). Its *Riemann curvature tensor*

$$R_{\lambda\mu\beta}^{\alpha} = \partial_{\lambda}K_{\mu\beta}^{\alpha} - \partial_{\mu}K_{\lambda\beta}^{\alpha} + K_{\lambda\beta}^{\gamma}K_{\mu\gamma}^{\alpha} - K_{\mu}^{\gamma}{}_{\beta}K_{\lambda}^{\alpha}{}_{\gamma}$$

has the temporal component

$$R_{\lambda\mu}{}^0{}_{\beta} = 0. \quad (5.93)$$

Then the equation for a Jacobi vector-field u along a geodesic c reads

$$\dot{q}^{\beta}\dot{q}^{\mu}(\nabla_{\beta}(\nabla_{\mu}u^{\alpha}) - R_{\lambda\mu\beta}^{\alpha}u^{\lambda}) = 0, \quad \nabla_{\beta}\dot{q}^{\alpha} = 0, \quad (5.94)$$

where ∇_{μ} denotes the covariant derivative relative to the connection \tilde{K} (see [Kobayashi and Nomizu (1963/9)]). Due to the relation (5.93), the equation (5.94) for the temporal component u^0 of a Jacobi field takes the form

$$\dot{q}^{\beta}\dot{q}^{\mu}(\partial_{\mu}\partial_{\beta}u^0 + K_{\mu}^{\gamma}{}_{\beta}\partial_{\gamma}u^0) = 0.$$

We chose its solution

$$u^0 = 0, \quad (5.95)$$

because all geodesics obey the constraint $\dot{t} = 0$.

Note that, in the case of a quadratic Lagrangian L , the equation (5.94) coincides with the Jacobi equation

$$w^j d_0(\partial_j \dot{\partial}_i L) + d_0(\dot{w}^j \dot{\partial}_i \dot{\partial}_j L) - w^j \partial_i \partial_j L = 0$$

for a Jacobi field on solutions of the Lagrangian equations for L . This equation is the Lagrangian equation for the vertical extension L_V of the Lagrangian L (see [Dittrich and Reuter (1994); Mangiarotti and Sardanashvily (1998); Mangiarotti *et. al.* (1999)]).

Let us consider a quadratic Newtonian system with a Riemannian inertia tensor μ_{ij} . Given a reference frame $(t, q^i) \equiv q^{\alpha}$, this inertia tensor is extended to the following Riemannian metric on Q

$$\bar{g}_{00} = 1, \quad \bar{g}_{0i} = 0, \quad \bar{g}_{ij} = \mu_{ij}.$$

However, its covariant derivative with respect to the connection \tilde{K} (5.83) does not vanish in general. Nevertheless, due to the relations (5.91) and

(5.95), we get the well-known formula for a Jacobi vector-field u along a geodesic c :

$$\int_a^b (\bar{g}_{\lambda\mu}(\dot{q}^\alpha \nabla_\alpha u^\lambda)(\dot{q}^\beta \nabla_\beta u^\mu) + R_{\lambda\mu\alpha\nu} u^\lambda u^\alpha \dot{q}^\mu \dot{q}^\nu) dt \\ + \bar{g}_{\lambda\mu} \dot{q}^\alpha \nabla_\alpha u^\lambda u'^\mu|_{t=a} - \bar{g}_{\lambda\mu} \dot{q}^\alpha \nabla_\alpha u^\lambda u'^\mu|_{t=b} = 0.$$

Therefore, the following assertions also remain true [Kobayashi and Nomizu (1963/9)]: (i) if the sectional curvature $R_{\lambda\mu\alpha\nu} u^\lambda u^\alpha \dot{q}^\mu \dot{q}^\nu$ is positive on a geodesic c , this geodesic has no conjugate points; (ii) if the sectional curvature $R_{\lambda\mu\alpha\nu} u^\lambda u^\alpha v^\mu v^\nu$, where u and v are arbitrary unit vectors on a Riemannian manifold Q less than $k < 0$, then, for every geodesic, the distance between two consecutive conjugate points is at most π/\sqrt{k} .

For example, let us consider a 1D motion described by the Lagrangian

$$L = \frac{1}{2}(\dot{q}^1)^2 - \phi(q^1),$$

where ϕ is a potential. The corresponding Lagrangian equation is equivalent to the geodesic one on the 2D Euclidean space \mathbb{R}^2 with respect to the connection \tilde{K} whose non-zero component is $\tilde{K}_0^1{}_0 = -\partial_1\phi$. The curvature of \tilde{K} has the non-zero component

$$R_{10}^1{}_0 = \partial_1 \tilde{K}_0^1{}_0 = -\partial_1^2 \phi.$$

Choosing the particular Riemannian metric g with components

$$\bar{g}_{11} = 1, \quad \bar{g}_{01} = 0, \quad \bar{g}_{00} = 1,$$

we come to the formula

$$\int_a^b [(\dot{q}^\mu \partial_\mu u^1)^2 - \partial_1^2 \phi (u^1)^2] dt = 0$$

for a Jacobi vector-field u , which vanishes at points a and b . Then we get that, if $\partial_1^2 \phi < 0$ at points of c , this motion has no conjugate points. In particular, let us consider the oscillator $\phi = k(q^1)^2/2$. In this case, the sectional curvature is $R_{0101} = -k$, while the half-period of this oscillator is exactly π/\sqrt{k} .

5.6.6 Constraints

Recall that symplectic and Poisson manifolds give an adequate Hamiltonian formulation of classical and quantum conservative mechanics. This

is also the case of *presymplectic Hamiltonian systems*. Recall that every presymplectic form can be represented as a pull-back of a symplectic form by a coisotropic imbedding (see e.g., [Gotay (1982); Mangiarotti and Sardanashvily (1998)]), a presymplectic Hamiltonian systems can be seen as *Dirac constraint systems* [Cariñena *et al.* (1995); Mangiarotti and Sardanashvily (1998)]. An autonomous Lagrangian system also exemplifies a presymplectic Hamiltonian system where a presymplectic form is the exterior differential of the Poincaré–Cartan form, while a Hamiltonian is the energy function [Cariñena and Rañada (1993); León *et al.* (1996); Mangiarotti and Sardanashvily (1998); Muñoz and Román (1992)]. A generic example of conservative Hamiltonian mechanics is a regular Poisson manifold (Z, w) where a Hamiltonian is a real function \mathcal{H} on Z . Given the corresponding Hamiltonian vector-field $\vartheta_{\mathcal{H}} = w^{\#}(d\mathcal{H})$, the closed subbundle $\vartheta_{\mathcal{H}}(Z)$ of the tangent bundle TZ is an autonomous first-order dynamical equation on a manifold Z , called the Hamiltonian equations. The evolution equation on the Poisson algebra $C^{\infty}(Z)$ is the *Lie derivative* $\mathfrak{L}_{\vartheta_{\mathcal{H}}}f = \{\mathcal{H}, f\}$, expressed into the *Poisson bracket* of the Hamiltonian \mathcal{H} and functions f on Z . However, this description cannot be extended in a straightforward manner to time-dependent mechanics subject to time-dependent transformations.

The existent formulations of time-dependent mechanics imply usually a preliminary splitting of a configuration space $Q = \mathbb{R} \times M$ and a *momentum phase-space manifold* $\Pi = \mathbb{R} \times Z$, where Z is a Poisson manifold [Cariñena and F.Núñez (1993); Chinea *et al.* (1994); Echeverría *et al.* (1991); Hamoui and Lichnerowicz (1984); Morandi *et al.* (1990); León and Marrero (1993)]. From the physical viewpoint, this means that a certain reference frame is chosen. In this case, the momentum phase-space Π is with the Poisson product of the zero Poisson structure on \mathbb{R} and the Poisson structure on Z . A Hamiltonian is defined as a real function \mathcal{H} on Π . The corresponding Hamiltonian vector-field $\vartheta_{\mathcal{H}}$ on Π is vertical with respect to the fibration $\Pi \rightarrow \mathbb{R}$. Due to the canonical imbedding

$$\Pi \times T\mathbb{R} \rightarrow T\Pi, \quad (5.96)$$

one introduces the vector-field

$$\gamma_{\mathcal{H}} = \partial_t + \vartheta_{\mathcal{H}}, \quad (5.97)$$

where ∂_t is the standard vector-field on \mathbb{R} [Hamoui and Lichnerowicz (1984)]. The first-order dynamical equation $\gamma_{\mathcal{H}}(\Pi) \subset T\Pi$ on the mani-

fold Π plays the role of Hamiltonian equations. The evolution equation on the Poisson algebra $C^\infty(\Pi)$ is given by the *Lie derivative*

$$\mathfrak{L}_{\gamma_{\mathcal{H}}} f = \partial_t f + \{\mathcal{H}, f\}.$$

This is not the case of mechanical systems subject to time-dependent transformations. These transformations, including canonical and inertial frame transformations, violate the splitting $\mathbb{R} \times Z$. As a consequence, there is no canonical imbedding (5.96), and the vector-field (5.97) is not well defined. At the same time, one can treat the imbedding (5.96) as a trivial connection on the bundle $\Pi \rightarrow \mathbb{R}$, while $\gamma_{\mathcal{H}}$ (5.97) is the sum of the horizontal lift onto Π of the vector-field ∂_t by this connection and of the vertical vector-field $\vartheta_{\mathcal{H}}$.

Let $Q \rightarrow \mathbb{R}$ be a fibre bundle coordinated by (t, q^i) , and $J^1(\mathbb{R}, Q)$ its 1-jet space, equipped with the adapted coordinates (t, q^i, q_t^i) . Recall that there is a canonical imbedding λ given by (5.64) onto the affine subbundle of $TQ \rightarrow Q$ of elements $v \in TQ$ such that $v|dt = 1$. This subbundle is modelled over the vertical tangent bundle $VQ \rightarrow Q$. As a consequence, there is a 1–1 correspondence between the connections Γ on the fibre bundle $Q \rightarrow \mathbb{R}$, treated as sections of the affine jet bundle $\pi_0^1 : J^1(\mathbb{R}, Q) \rightarrow Q$ [Mangiarotti *et. al.* (1999)], and the nowhere vanishing vector-fields $\Gamma = \partial_t + \Gamma^i \partial_i$ on Q , called horizontal vector-fields, such that $\Gamma|dt = 1$ [Mangiarotti and Sardanashvily (1998); Mangiarotti *et. al.* (1999)]. The corresponding *covariant differential* reads

$$D_\Gamma = \lambda - \Gamma : J^1(\mathbb{R}, Q) \longrightarrow VQ, \quad q^i \circ D_\Gamma = q_t^i - \Gamma^i.$$

Let us also recall the total derivative $d_t = \partial_t + q_t^i \partial_i + \cdots$ and the exterior algebra homomorphism

$$h_0 : \phi dt + \phi_i dy^i \mapsto (\phi + \phi_i q_t^i) dt \quad (5.98)$$

which sends exterior forms on $Q \rightarrow \mathbb{R}$ onto the horizontal forms on $J^1(\mathbb{R}, Q) \rightarrow \mathbb{R}$, and vanishes on contact forms $\theta^i = dy^i - q_t^i dt$.

Lagrangian time-dependent mechanics follows directly Lagrangian field theory (see [Giachetta (1992); Krupkova (1997); León *et. al.* (1997); Mangiarotti and Sardanashvily (1998); Massa and Pagani (1994)], as well as subsection 5.9 below). This means that we have a configuration space $Q \rightarrow \mathbb{R}$ of a mechanical system, and a Lagrangian is defined as a *horizontal density* on the velocity phase-space $J^1(\mathbb{R}, Q)$,

$$L = \mathcal{L} dt, \quad \text{with} \quad \mathcal{L} : J^1(\mathbb{R}, Q) \rightarrow \mathbb{R}. \quad (5.99)$$

A generic *momentum phase-space manifold* of time-dependent mechanics is a fibre bundle $\Pi \rightarrow \mathbb{R}$ with a regular Poisson structure whose characteristic distribution belongs to the vertical tangent bundle $V\Pi$ of $\Pi \rightarrow \mathbb{R}$ [Hamoui and Lichnerowicz (1984)]. However, such a Poisson structure cannot give dynamical equations. A first-order dynamical equation on $\Pi \rightarrow \mathbb{R}$, by definition, is a section of the affine jet bundle $J^1(\mathbb{R}, \Pi) \rightarrow \Pi$, i.e., a connection on $\Pi \rightarrow \mathbb{R}$. Being a horizontal vector-field, such a connection cannot be a Hamiltonian vector-field with respect to the above Poisson structure on Π .

One can overcome this difficulty as follows. Let $Q \rightarrow \mathbb{R}$ be a configuration bundle of time-dependent mechanics. The corresponding momentum phase-space is the vertical cotangent bundle $\Pi = V^*Q \rightarrow \mathbb{R}$, called the Legendre bundle, while the cotangent bundle T^*Q is the homogeneous momentum phase-space. T^*Q admits the canonical Liouville form Ξ and the symplectic form $d\Xi$, together with the corresponding non-degenerate Poisson bracket $\{, \}_T$ on the ring $C^\infty(T^*Q)$. Let us consider the subring of $C^\infty(T^*Q)$ which comprises the pull-backs ζ^*f onto T^*Q of functions f on the vertical cotangent bundle V^*Q by the canonical fibration

$$\zeta : T^*Q \rightarrow V^*Q. \quad (5.100)$$

This subring is closed under the Poisson bracket $\{, \}_T$, and V^*Q admits the regular Poisson structure $\{, \}_V$ such that [Vaisman (1994)]

$$\zeta^*\{f, g\}_V = \{\zeta^*f, \zeta^*g\}_T.$$

Its characteristic distribution coincides with the vertical tangent bundle VV^*Q of $V^*Q \rightarrow \mathbb{R}$. Given a section h of the bundle (5.100), let us consider the pull-back forms

$$\Theta = h^*(\Xi \wedge dt), \quad \Omega = h^*(d\Xi \wedge dt) \quad (5.101)$$

on V^*Q , but these forms are independent of a section h and are canonical exterior forms on V^*Q . The pull-backs $h^*\Xi$ are called the Hamiltonian forms. With Ω , the Hamiltonian vector-field ϑ_f for a function f on V^*Q is given by the relation

$$\vartheta_f \rfloor \Omega = -df \wedge dt,$$

while the Poisson bracket (5.6.6) is written as

$$\{f, g\}_V dt = \vartheta_g \rfloor \vartheta_f \rfloor \Omega.$$

The pair (V^*Q, Ω) is the particular *polysymplectic phase-space* of the covariant Hamiltonian field theory (see [Cariñena *et. al.* (1991); Giachetta *et. al.* (1997); Gotay (1991a); Sardanashvily (1995)] for a survey). Following its general scheme, we can formulate the Hamiltonian time-dependent mechanics as follows [Mangiarotti and Sardanashvily (1998); Sardanashvily (1998)].

Recall that connection γ on the Legendre bundle $V^*Q \rightarrow \mathbb{R}$ is called *canonical* if the corresponding horizontal vector-field is canonical for the Poisson structure on V^*Q , i.e., the form $\gamma \rfloor \Omega$ is closed. Such a form is necessarily exact. A canonical connection γ is said to be a *Hamiltonian connection* if

$$\gamma \rfloor \Omega = dH, \quad (5.102)$$

where H is a *Hamiltonian form* on V^*Q . Every Hamiltonian form admits a unique Hamiltonian connection γ_H , while any canonical connection is locally a Hamiltonian one. Given a Hamiltonian form H , the kernel of the covariant differential D_{γ_H} , associated with the Hamiltonian connection γ_H , is a closed imbedded subbundle of the jet bundle $J^1(\mathbb{R}, V^*Q) \rightarrow \mathbb{R}$, and so is the system of first-order PDEs on the Legendre bundle $V^*Q \rightarrow \mathbb{R}$. These are the *Hamiltonian equations* in time-dependent mechanics, while the *Lie derivative*

$$\mathfrak{L}_{\gamma_H} f = \gamma_H \rfloor df \quad (5.103)$$

defines the evolution equation on $C^\infty(V^*Z)$. This Hamiltonian dynamics is equivalent to the Lagrangian one for hyperregular Lagrangians, while a degenerate Lagrangian involves a set of associated Hamiltonian forms in order to exhaust solutions of the Lagrangian equations [Giachetta *et. al.* (1997); Sardanashvily (1994); Sardanashvily (1995)].

Since γ_H is not a vertical vector-field, the r.h.s. of the evolution equation (5.103) is not expressed into the Poisson bracket in a canonical way, but contains a frame-dependent term. Every connection Γ on the configuration bundle $Q \rightarrow \mathbb{R}$ is an affine section of the bundle (5.100), and defines the Hamiltonian form $H_\Gamma = \Gamma^* \Xi$ on V^*Q . The corresponding Hamiltonian connection is the canonical lift $V^*\Gamma$ of Γ onto the Legendre bundle V^*Q [Giachetta *et. al.* (1997); Mangiarotti *et. al.* (1999)]. Then any Hamiltonian form H on V^*Q admits splittings

$$H = H_\Gamma - \tilde{\mathcal{H}}_\Gamma dt, \quad \text{with} \quad \gamma_H = V^*\Gamma + \vartheta_{\tilde{\mathcal{H}}_\Gamma}, \quad (5.104)$$

where $\vartheta_{\tilde{\mathcal{H}}_\Gamma}$ is the vertical Hamiltonian field for the function $\tilde{\mathcal{H}}_\Gamma$, which the energy function with respect to the reference frame Γ . With the splitting (5.104), the evolution equation (5.103) takes the form

$$\mathfrak{L}_{\gamma_H} f = V^* \Gamma \rfloor H + \{\tilde{\mathcal{H}}_\Gamma, f\}_V. \quad (5.105)$$

Let the configuration bundle $Q \rightarrow \mathbb{R}$ with an m D typical fibre M be co-ordinated by (t, q^i) . Then Legendre bundle V^*Q and the cotangent bundle T^*Q admit holonomic coordinates $(t, q^i, p_i = \dot{q}_i)$ and (t, q^i, p_i, p) , respectively. Relative to these coordinates, a *Hamiltonian form* H on V^*Q reads

$$H = h^* \Xi = p_i dq^i - \mathcal{H} dt. \quad (5.106)$$

It is the well-known integral *invariant of Poincaré–Cartan*, where \mathcal{H} is a Hamiltonian in time-dependent mechanics. The expression (5.106) shows that \mathcal{H} fails to be a scalar under time-dependent transformations. Therefore, the evolution equation (5.105) takes the local form

$$\mathfrak{L}_{\gamma_H} f = \partial_t f + \{\mathcal{H}, f\}_V, \quad (5.107)$$

but one should bear in mind that the terms in its r.h.s., taken separately, are not well-behaved objects under time-dependent transformations. In particular, the equality $\{\mathcal{H}, f\}_V = 0$ is not preserved under time-dependent transformations.

Every Lagrangian L defines the *Legendre map*

$$\hat{L} : J^1(\mathbb{R}, Q) \longrightarrow V^*Q, \quad \text{locally given by} \quad p_i \circ \hat{L} = \pi_i, \quad (5.108)$$

whose image $N_L = \hat{L}(J^1(\mathbb{R}, Q)) \subset V^*Q$ is called the *Lagrangian constraint space*. We state the comprehensive relationship between solutions of the Lagrangian equations for an almost regular Lagrangian L and solutions in N_L of the Hamiltonian equations for associated Hamiltonian forms.

5.6.7 Time-Dependent Lagrangian Dynamics

Given a Lagrangian L on the velocity phase-space $J^1(\mathbb{R}, Q)$, we follow the *first variational formula* of [Giachetta *et. al.* (1997); Mangiarotti and Sardanashvily (1998); Sardanashvily (1997)], which gives the canonical decomposition of the Lie derivative

$$\mathfrak{L}_{j^1 u} L = (j^1 u \rfloor \mathcal{L}) dt \quad (5.109)$$

of L along a projectable vector-field u on $Q \rightarrow \mathbb{R}$. We have

$$j^1 u \rfloor \mathcal{L} = u_V \rfloor \mathcal{E}_L + d_t(u \rfloor H_L),$$

where $u_V = (u \rfloor \theta^i) \partial_i$,

$$H_L = L + \pi_i \theta^i, \quad \pi_i = \partial_i^t \mathcal{L}, \quad (5.110)$$

is the *Poincaré–Cartan form* and

$$\mathcal{E}_L : J^2(\mathbb{R}, Q) \rightarrow V^*Q, \quad \mathcal{E}_L = (\partial_i - d_t \pi_i) \mathcal{L} \bar{d}q^i$$

is the *Euler–Lagrangian operator* associated with L . The kernel $\text{Ker } \mathcal{E}_L \subset J^2(\mathbb{R}, Q)$ of \mathcal{E}_L defines the Lagrangian equations on Q , given by the coordinate relations

$$(\partial_i - d_t \pi_i) \mathcal{L} = 0. \quad (5.111)$$

On-shell, the first variational formula (5.109) leads to the weak identity

$$\mathfrak{L}_{j^1 u} L \approx d_t(u \rfloor H_L) dt,$$

and then, if $\mathfrak{L}_{j^1 u} L = 0$, to the weak conservation law

$$d_t(u \rfloor H_L) = -d_t \mathcal{J} \approx 0$$

of the symmetry current \mathcal{J} , given by

$$\mathcal{J} = -(u \rfloor H_L) = -\pi_i (u^t \dot{q}^i - u^i) - u^t \mathcal{L}.$$

Being the *Lepagean equivalent* of the Lagrangian L on $J^1(\mathbb{R}, Q)$ (i.e., $L = h_0(H_L)$ where h_0 is the map (5.98), see [Giachetta *et. al.* (1997); Mangiarotti and Sardanashvily (1998); Sardanashvily (1997)]), the *Poincaré–Cartan form* H_L (5.110) is also the Lepagean equivalent of the Lagrangian on the repeated jet space $J^1(\mathbb{R}, J^1(\mathbb{R}, Q))$,

$$\bar{L} = \hat{h}_0(H_L) = (\mathcal{L} + (\hat{q}^i - \dot{q}^i) \pi_i) dt, \quad \hat{h}_0(dy^i) = \hat{q}_t^i dt,$$

coordinated by $(t, q^i, \dot{q}^i, \hat{q}^i, \ddot{q}^i)$. The Euler–Lagrangian operator $\mathcal{E}_{\bar{L}} : J^1(\mathbb{R}, J^1(\mathbb{R}, Q)) \rightarrow V^*J^1(\mathbb{R}, Q)$ for \bar{L} is locally given by

$$\begin{aligned} \mathcal{E}_{\bar{L}} &= (\partial_i \mathcal{L} - \hat{d}_t \pi_i + \partial_i \pi_j (\hat{q}^j - \dot{q}^j)) \bar{d}q^i + \partial_i^t \pi_j (\hat{q}^j - \dot{q}^j) \bar{d}\dot{q}^i, \quad (5.112) \\ \text{with} \quad \hat{d}_t &= \partial_t + \hat{q}^i \partial_i + \ddot{q}^i \partial_i^t. \end{aligned}$$

Its kernel $\text{Ker } \mathcal{E}_{\bar{L}} \subset J^1(\mathbb{R}, J^1(\mathbb{R}, Q))$ defines the *Cartan equations*

$$\partial_i^t \pi_j (\hat{q}^j - \dot{q}^j) = 0, \quad \partial_i \mathcal{L} - \hat{d}_t \pi_i + (\hat{q}^j - \dot{q}^j) \partial_i \pi_j = 0. \quad (5.113)$$

Since $\mathcal{E}_{\bar{L}}|_{J^2(\mathbb{R}, Q)} = \mathcal{E}_L$, the Cartan equations (5.113) are equivalent to the Lagrangian equations (5.111) on integrable sections $\bar{c} = \dot{c}$ of $J^1(\mathbb{R}, Q) \rightarrow \mathbb{R}$. These equations are equivalent in the case of regular Lagrangians.

On sections $\bar{c} : \mathbb{R} \rightarrow J^1(\mathbb{R}, Q)$, the Cartan equations (5.113) are equivalent to the relation

$$\bar{c}^*(u \rfloor dH_L) = 0, \quad (5.114)$$

which is assumed to hold for all vertical vector-fields u on $J^1(\mathbb{R}, Q) \rightarrow \mathbb{R}$.

With the Poincaré–Cartan form H_L (5.110), we have the *Legendre map*

$$\hat{H}_L : J^1(\mathbb{R}, Q) \longrightarrow T^*Q, \quad (p_i, p) \circ \hat{H}_L = (\pi_i, \mathcal{L} - \pi_i \dot{q}^i).$$

Let $Z_L = \hat{H}_L(J^1(\mathbb{R}, Q))$ be an imbedded subbundle $i_L : Z_L \hookrightarrow T^*Q$ of $T^*Q \rightarrow Q$. It admits the pull-back de Donder form $i_L^* \Xi$. We have

$$H_L = \hat{H}_L^* \Xi_L = \hat{H}_L^*(i_L^* \Xi).$$

By analogy with the Cartan equations (5.114), the *Hamilton–de Donder equations* for sections \bar{r} of $T^*Q \rightarrow \mathbb{R}$ are written as

$$\bar{r}^*(u \rfloor d\Xi_L) = 0 \quad (5.115)$$

where u is an arbitrary vertical vector-field on $T^*Q \rightarrow \mathbb{R}$ [Lopez and Marsden (2003)].

Let the Legendre map $\hat{H}_L : J^1(\mathbb{R}, Q) \rightarrow Z_L$ be a submersion. Then a section \bar{c} of $J^1(\mathbb{R}, Q) \rightarrow \mathbb{R}$ is a solution of the Cartan equations (5.114) iff $\hat{H}_L \circ \bar{c}$ is a solution of the Hamilton–de Donder equations (5.115), i.e., Cartan and Hamilton–de Donder equations are quasi-equivalent [Gotay (1991a); Lopez and Marsden (2003)].

5.6.8 Time-Dependent Hamiltonian Dynamics

Let the Legendre bundle $V^*Q \rightarrow \mathbb{R}$ be provided with the holonomic coordinates (t, q^i, \dot{q}^i) . Relative to these coordinates, the *canonical 3-form* Ω (5.101) and the *canonical Poisson structure* on V^*Q read

$$\Omega = dp_i \wedge dq^i \wedge dt, \quad (5.116)$$

$$\{f, g\}_V = \partial^i f \partial_i g - \partial^i g \partial_i f, \quad (f, g \in C^\infty V^*Q). \quad (5.117)$$

The corresponding *symplectic foliation* coincides with the fibration $V^*Q \rightarrow \mathbb{R}$. The symplectic forms on the fibres of $V^*Q \rightarrow \mathbb{R}$ are the pull-backs $\Omega_t = dp_i \wedge dq^i$ of the canonical symplectic form on the typical fibre

T^*M of the Legendre bundle $V^*Q \rightarrow \mathbb{R}$ with respect to trivialization maps [Cariñena and Rañada (1989); Hamoui and Lichnerowicz (1984); Sardanashvily (1998)]. Given such a trivialization, the Poisson structure (5.117) is isomorphic to the product of the zero Poisson structure on \mathbb{R} and the canonical symplectic structure on T^*M .

An automorphism ρ of the Legendre bundle $V^*Q \rightarrow \mathbb{R}$ is a canonical transformation of the Poisson structure (5.117) iff it preserves the canonical 3-form Ω (5.116). Let us emphasize that canonical transformations are compatible with the fibration $V^*Q \rightarrow \mathbb{R}$, but not necessarily with the fibration $\pi_Q : V^*Q \rightarrow Q$.

With respect to the Poisson bracket (5.117), the Hamiltonian vector-field ϑ_f for a function f on the *momentum phase-space manifold* V^*Q is given by

$$\vartheta_f = \partial^i f \partial_i - \partial_i f \partial^i.$$

A Hamiltonian vector-field, by definition, is canonical. Conversely, every vertical canonical vector-field on the Legendre bundle $V^*Q \rightarrow \mathbb{R}$ is locally a Hamiltonian vector-field.

To prove this, let σ be a one-form on V^*Q . If $\sigma \wedge dt$ is closed form, it is exact. Since V^*Q is diffeomorphic to $\mathbb{R} \times T^*M$, we have the *de Rham cohomology group*

$$H^2(V^*Q) = H^0(\mathbb{R}) \otimes H^2(T^*M) \oplus H^1(\mathbb{R}) \otimes H^1(T^*M).$$

The form $\sigma \wedge dt$ belongs to its second item which is zero.

If the two-form $\sigma \wedge dt$ is exact, then $\sigma \wedge dt = dg \wedge dt$ locally [Giachetta *et. al.* (1997)].

Let $\gamma = \partial_t + \gamma^i \partial_i + \gamma_i \partial^i$ be a canonical connection on the Legendre bundle $V^*Q \rightarrow \mathbb{R}$. Its components obey the relations

$$\partial^i \gamma^j - \partial^j \gamma^i = 0, \quad \partial_i \gamma_j - \partial_j \gamma_i = 0, \quad \partial_j \gamma^i + \partial^i \gamma_j = 0.$$

Canonical connections constitute an affine space modelled over the vector space of vertical canonical vector-fields on $V^*Q \rightarrow \mathbb{R}$.

If γ is a canonical connection, then the form $\gamma \rfloor \Omega$ is exact. Every connection Γ on $Q \rightarrow \mathbb{R}$ induces the connection on $V^*Q \rightarrow \mathbb{R}$,

$$V^*\Gamma = \partial_t + \Gamma^i \partial_i - p_i \partial_j \Gamma^i \partial^j,$$

which is a *Hamiltonian connection* for the frame *Hamiltonian form*

$$V^*\Gamma \rfloor \Omega = dH_\Gamma, \quad H_\Gamma = p_i dq^i - p_i \Gamma^i dt. \quad (5.118)$$

Thus, every canonical connection γ on V^*Q defines an exterior one-form H modulo closed forms so that $dH = \gamma \rfloor \Omega$. Such a form is called a *locally Hamiltonian form*.

Every locally Hamiltonian form on the momentum phase-space V^*Q is locally a Hamiltonian form modulo closed forms. Given locally Hamiltonian forms H_γ and $H_{\gamma'}$, their difference $\sigma = H_\gamma - H_{\gamma'}$ is a one-form on V^*Q such that the two-form $\sigma \wedge dt$ is closed. The form $\sigma \wedge dt$ is exact and $\sigma = fdt + dg$ locally. Put $H_{\gamma'} = H_\Gamma$ where Γ is a connection on $V^*Q \rightarrow \mathbb{R}$. Then H_γ modulo closed forms takes the local form $H_\gamma = H_\Gamma + fdt$, and coincides with the pull-back of the Liouville form Ξ on T^*Q by the local section $p = -p_i\Gamma^i + f$ of the fibre bundle (5.100).

Conversely, each Hamiltonian form H on the momentum phase-space V^*Q admits a unique canonical connection γ_H on $V^*Q \rightarrow \mathbb{R}$ such that the relation (5.102) holds. Given a Hamiltonian form H , its exterior differential

$$dH = h^*d\Xi = (dp_i + \partial_i H dt) \wedge (dq^i - \partial^i H dt)$$

is a presymplectic form of constant rank $2m$ since the form

$$(dH)^m = (dp_i \wedge dq^i)^m - m(dp_i \wedge dq^i)^{m-1} \wedge dH \wedge dt$$

is nowhere vanishing. It is also seen that $(dH)^m \wedge dt \neq 0$. It follows that the kernel of dH is a 1D distribution. Then the desired *Hamiltonian connection*

$$\gamma_H = \partial_t + \partial^i H \partial_i - \partial_i H \partial^i \quad (5.119)$$

is a unique vector-field γ_H on V^*Q such that $\gamma_H \rfloor dH = 0$ and $\gamma_H \rfloor dt = 1$.

Hamiltonian forms constitute an affine space modelled over the vector space of horizontal densities fdt on $V^*Q \rightarrow \mathbb{R}$, i.e., over $C^\infty(V^*Q)$. Therefore Hamiltonian connections γ_H form an affine space modelled over the vector space of Hamiltonian vector-fields. Every Hamiltonian form H defines the associated Hamiltonian map

$$\widehat{H} = j^1\pi_Q \circ \gamma_H : \partial_t + \partial^i \mathcal{H} : V^*Q \rightarrow J^1(\mathbb{R}, Q). \quad (5.120)$$

With the Hamiltonian map (5.120), we have another Hamiltonian form

$$H_{\widehat{H}} = -\widehat{H} \rfloor \Theta = p_i dq^i - p_i \partial^i \mathcal{H}.$$

Note that $H_{\widehat{H}} = H$ iff H is a frame Hamiltonian form.

Given a Hamiltonian connection γ_H (5.119), the corresponding Hamiltonian equations $D_{\gamma_H} = 0$ take the coordinate form

$$\dot{q}^i = \partial^i \mathcal{H}, \quad \dot{p}_i = -\partial_i \mathcal{H}. \quad (5.121)$$

Their classical solutions are integral sections of the Hamiltonian connection γ_H , i.e., $\dot{r} = \gamma_H \circ r$. On sections r of the Legendre bundle $V^*Q \rightarrow \mathbb{R}$, the Hamiltonian equations (5.121) are equivalent to the relation

$$r^*(u \rfloor dH) = 0 \quad (5.122)$$

which is assumed to hold for any vertical vector-field u on $V^*Q \rightarrow \mathbb{R}$.

A Hamiltonian form H (5.118) is the Poincaré–Cartan form for the Lagrangian on the jet space $J^1(\mathbb{R}, V)^*Q$,

$$L_H = h_0(H) = (p_i \dot{q}^i - \mathcal{H})\omega. \quad (5.123)$$

Given a projectable vector-field u on the configuration bundle $Q \rightarrow \mathbb{R}$ and its lift onto the Legendre bundle $V^*Q \rightarrow \mathbb{R}$,

$$\begin{aligned} \tilde{u} &= u^t \partial_t + u^i \partial_i - \partial_i u^j p_j \partial^i, & \text{we have} \\ \mathfrak{L}_{\tilde{u}} H &= \mathfrak{L}_{J^1 \tilde{u}} L_H. \end{aligned} \quad (5.124)$$

Note that the Hamiltonian equations (5.121) for H are exactly the Lagrangian equations for L_H , i.e., they characterize the kernel of the Euler–Lagrangian operator

$$\mathcal{E}_H : J^1(\mathbb{R}, V)^*Q \rightarrow V^*V^*Q, \quad \mathcal{E}_H = (\dot{q}^i - \partial^i \mathcal{H}) \bar{d}p_i - (\dot{p}_i + \partial_i \mathcal{H}) \bar{d}q^i$$

for the Lagrangian L_H , called the Hamiltonian operator for H .

Using the relation (5.124), let us get the *Hamiltonian conservation laws* in time-dependent mechanics. As in field theory, by gauge transformations in time-dependent mechanics are meant automorphism of the configuration bundle $Q \rightarrow \mathbb{R}$, but only over translations of the base \mathbb{R} . Then, projectable vector-fields on $V^*Q \rightarrow \mathbb{R}$,

$$u = u^t \partial_t + u^i \partial_i, \quad u \rfloor dt = u^t = \text{const}, \quad (5.125)$$

can be seen as generators of local 1-parameter groups of local gauge transformations. Given a Hamiltonian form H (5.165), its Lie derivative (5.124) reads

$$\mathfrak{L}_{\tilde{u}} H = \mathfrak{L}_{J^1 \tilde{u}} L_H = (-u^t \partial_t \mathcal{H} + p_i \partial_t u^i - u^i \partial_i \mathcal{H} + \partial_j u^i p_i \partial^j \mathcal{H}) dt. \quad (5.126)$$

The first variational formula (5.109) applied to the Lagrangian L_H (5.123) leads to the weak identity $\mathfrak{L}_{\tilde{u}} H \approx d_t(u \rfloor H) dt$. If the Lie derivative (5.126) vanishes, we have the *conserved symmetry current*

$$\mathcal{J}_u = u \rfloor dH = p_i u^i - u^t \mathcal{H}, \quad (5.127)$$

along u . Every vector-field (5.125) is a superposition of a vertical vector-field and a reference frame on $Q \rightarrow \mathbb{R}$. If u is a vertical vector-field, \mathcal{J}_u is the Nöther current

$$\mathcal{J}_u(q) = u \rfloor q = p_i u^i, \quad (q = p_i \bar{d}q^i \in V^*Q). \quad (5.128)$$

The symmetry current along a reference frame Γ , given by

$$\mathcal{J}_\Gamma = p_i \Gamma^i - \mathcal{H} = -\tilde{\mathcal{H}}_\Gamma,$$

is the energy function with respect to the reference frame Γ , taken with the sign minus [Echeverría *et. al.* (1995); Mangiarotti and Sardanashvily (1998); Sardanashvily (1998)]. Given a Hamiltonian form H , the energy functions $\tilde{\mathcal{H}}_\Gamma$ constitute an affine space modelled over the vector space of Nöther currents. Also, given a Hamiltonian form H , the conserved currents (5.127) form a Lie algebra with respect to the Poisson bracket

$$\{\mathcal{J}_u, \mathcal{J}_{u'}\}_V = \mathcal{J}_{[u, u']}.$$

The second of the above constructions enables us to represent the r.h.s. of the evolution equation (5.107) as a pure Poisson bracket. Given a Hamiltonian form $H = h^* \Xi$, let us consider its pull-back $\zeta^* H$ onto the cotangent bundle T^*Q . Note that the difference $\Xi - \zeta^* H$ is a horizontal one-form on $T^*Q \rightarrow \mathbb{R}$, while

$$\mathcal{H}^* = \partial_t \rfloor (\Xi - \zeta^* H) = p + \mathcal{H} \quad (5.129)$$

is a function on T^*Q . Then the relation

$$\zeta^* (\mathfrak{L}_{\gamma_H} f) = \{\mathcal{H}^*, \zeta^* f\}_T \quad (5.130)$$

holds for every function $f \in C^\infty(V^*Q)$. In particular, given a projectable vector-field u (5.125), the symmetry current \mathcal{J}_u (5.127) is conserved iff

$$\{\mathcal{H}^*, \zeta^* \mathcal{J}_u\}_T = 0.$$

Moreover, let $\vartheta_{\mathcal{H}^*}$ be the Hamiltonian vector-field for the function \mathcal{H}^* (5.129) with respect to the canonical Poisson structure $\{, \}_T$ on T^*Q . Then

$$T\zeta(\vartheta_{\mathcal{H}^*}) = \gamma_H.$$

5.6.9 Time-Dependent Constraints

The relation (5.130) enables us to extend the constraint algorithm of conservative mechanics and time-dependent mechanics on a product $\mathbb{R} \times M$ (see [Chinea *et. al.* (1994); León and Marrero (1993)]) to mechanical systems subject to time-dependent transformations.

Let H be a Hamiltonian form on the *momentum phase-space manifold* V^*Q . In accordance with the relation (5.130), a constraint $f \in I_N$ is preserved if the bracket in (5.130) vanishes. It follows that the solutions of the Hamiltonian equations (5.121) do not leave the constraint space N if

$$\{\mathcal{H}^*, \zeta^* I_N\}_T \subset \zeta^* I_N. \quad (5.131)$$

If the relation (5.131) fails to hold, let us introduce secondary constraints $\{\mathcal{H}^*, \zeta^* f\}_T$, $f \in I_N$, which belong to $\zeta^* C^\infty(V^*Q)$. If the collection of primary and secondary constraints is not closed with respect to the relation (5.131), let us add the tertiary constraints $\{\mathcal{H}^*, \{\mathcal{H}^*, \zeta^* f_a\}_T\}_T$ and so on.

Let us assume that N is a final constraint space for a Hamiltonian form H . If a Hamiltonian form H satisfies the relation (5.131), so is a Hamiltonian form

$$H_f = H - f dt, \quad (5.132)$$

where $f \in I'_N$ is a first class constraint. Though Hamiltonian forms H and H_f coincide with each other on the constraint space N , the corresponding Hamiltonian equations have different solutions on the constraint space N because $dH|_N \neq dH_f|_N$. At the same time, $d(i_N^* H) = d(i_N^* H_f)$. Therefore, let us introduce the constrained Hamiltonian form

$$H_N = i_N^* H_f, \quad (5.133)$$

which is the same for all $f \in I'_N$. Note that H_N (5.133) is not a true Hamiltonian form on $N \rightarrow \mathbb{R}$ in general. On sections r of the fibre bundle $N \rightarrow \mathbb{R}$, we can write the equations

$$r^*(u_N \rfloor dH_N) = 0, \quad (5.134)$$

where u_N is an arbitrary vertical vector-field on $N \rightarrow \mathbb{R}$. They are called the constrained Hamiltonian equations.

For any Hamiltonian form H_f (5.132), every solution of the Hamiltonian equations which lives in the constraint space N is a solution of the constrained Hamiltonian equations (5.134). The constrained Hamiltonian

equations can be written as

$$r^*(u_N \rfloor di_N^* H_f) = r^*(u_N \rfloor dH_f|_N) = 0. \quad (5.135)$$

They differ from the Hamiltonian equations (5.122) for H_f restricted to N which read

$$r^*(u \rfloor dH_f|_N) = 0, \quad (5.136)$$

where r is a section of $N \rightarrow \mathbb{R}$ and u is an arbitrary vertical vector-field on $V^*Q \rightarrow \mathbb{R}$. A solution r of the equations (5.136) satisfies the weaker condition (5.135).

One can also consider the problem of constructing a generalized Hamiltonian system, similar to that for Dirac constraint system in conservative mechanics [Mangiarotti and Sardanashvily (1998)]. Let H satisfies the condition $\{\mathcal{H}^*, \zeta^* I'_N\}_T \subset I_N$, whereas $\{\mathcal{H}^*, \zeta^* I'_N\}_T \not\subset I_N$. The goal is to find a constraint $f \in I_N$ such that the modified Hamiltonian $H - f dt$ would satisfy both the conditions

$$\{\mathcal{H}^* + \zeta^* f, \zeta^* I'_N\}_T \subset \zeta^* I_N, \quad \{\mathcal{H}^* + \zeta^* f, \zeta^* I_N\}_T \subset \zeta^* I_N.$$

The first of them is fulfilled for any $f \in I_N$, while the latter is an equation for a second-class constraint f .

Note that, in contrast with the conservative case, the Hamiltonian vector-fields ϑ_f for the first class constraints $f \in I'_N$ in time-dependent mechanics are not generators of gauge symmetries of a Hamiltonian form in general. At the same time, generators of gauge symmetries define an ideal of constraints as follows.

5.6.10 Lagrangian Constraints

Let us consider the Hamiltonian description of Lagrangian mechanical systems on a configuration bundle $Q \rightarrow \mathbb{R}$. If a Lagrangian is degenerate, we have the Lagrangian constraint subspace of the Legendre bundle V^*Q and a set of Hamiltonian forms associated with the same Lagrangian. Given a Lagrangian L (5.99) on the velocity phase-space $J^1(\mathbb{R}, Q)$, a Hamiltonian form H on the momentum phase-space V^*Q is said to be associated with L if H satisfies the relations

$$\widehat{L} \circ \widehat{H} \circ \widehat{L} = \widehat{L}, \quad \text{and} \quad H = H_{\widehat{H}} + \widehat{H}^* L \quad (5.137)$$

where \widehat{H} and \widehat{L} are the Hamiltonian map (5.120) and the Legendre map (5.108), respectively. Here, $\widehat{L} \circ \widehat{H}$ is the projector

$$p_i(z) = \pi_i(t, q^i, \partial^j \mathcal{H}(z)), \quad (z \in N_L), \quad (5.138)$$

from Π onto the Lagrangian constraint space $N_L = \widehat{L}(J^1(\mathbb{R}, Q))$. Therefore, $\widehat{H} \circ \widehat{L}$ is the projector from $J^1(\mathbb{R}, Q)$ onto $\widehat{H}(N_L)$. A Hamiltonian form is called weakly associated with a Lagrangian L if the condition (5.137) holds on the Lagrangian constraint space N_L .

If a bundle map $\Phi : V^*Q \rightarrow J^1(\mathbb{R}, Q)$ obeys the relation (5.137), then the Hamiltonian form $H = -\Phi] \Theta + \Phi^* L$ is weakly associated with the Lagrangian L . If $\Phi = \widehat{H}$, then H is associated with L [Giachetta *et. al.* (1997)].

Any Hamiltonian form H weakly associated with a Lagrangian L obeys the relation

$$H|_{N_L} = \widehat{H}^* H_L|_{N_L}, \quad (5.139)$$

where H_L is the Poincaré–Cartan form (5.110). The relation (5.137) takes the coordinate form

$$\mathcal{H}(z) = p_i \partial^i \mathcal{H} - \mathcal{L}(t, q^i, \partial^j \mathcal{H}(z)), \quad (z \in N_L). \quad (5.140)$$

Substituting (5.138) and (5.140) in (5.165), we get the relation (5.139).

The difference between *associated* and *weakly associated Hamiltonian forms* lies in the following. Let H be an associated Hamiltonian form, i.e., the equality (5.140) holds everywhere on V^*Q . The exterior differential of this equality leads to the relations

$$\begin{aligned} \partial_t \mathcal{H}(z) &= -(\partial_t \mathcal{L}) \circ \widehat{H}(z), & \partial_i \mathcal{H}(z) &= -(\partial_i \mathcal{L}) \circ \widehat{H}(z), \\ (p_i - (\partial_i^t \mathcal{L})(t, q^i, \partial_t^j \mathcal{H})) \partial_t^i \partial_t^a \mathcal{H} &= 0, & (z \in N_L). \end{aligned}$$

The last of them shows that the Hamiltonian form is not regular outside the Lagrangian constraint space N_L . In particular, any Hamiltonian form is weakly associated with the Lagrangian $L = 0$, while the associated Hamiltonian forms are only H_Γ .

Here we restrict our consideration to almost regular Lagrangians L , i.e., if: (i) the Lagrangian constraint space N_L is a closed imbedded subbundle $i_N : N_L \rightarrow V^*Q$ of the bundle $V^*Q \rightarrow Q$, (ii) the Legendre map $\widehat{L} : J^1(\mathbb{R}, Q) \rightarrow N_L$ is a fibre bundle, and (iii) the pre-image $\widehat{L}^{-1}(z)$ of any point $z \in N_L$ is a connected submanifold of $J^1(\mathbb{R}, Q)$.

A Hamiltonian form H weakly associated with an almost regular Lagrangian L exists iff the fibre bundle $J^1(\mathbb{R}, V)^*Q \rightarrow N_L$ admits a global section.

The condition (iii) leads to the following property [Giachetta *et. al.* (1997); Mangiarotti and Sardanashvily (1998)]. The Poincaré–Cartan form H_L for an almost regular Lagrangian L is constant on the connected pre-image $\widehat{L}^{-1}(z)$ of any point $z \in N_L$.

An immediate consequence of this fact is the following assertion [Giachetta *et. al.* (1997)]. All Hamiltonian forms weakly associated with an almost regular Lagrangian L coincide with each other on the Lagrangian constraint space N_L , and the Poincaré–Cartan form H_L for L is the pull-back

$$H_L = \widehat{L}^* H, \quad \pi_i \dot{q}^i - \mathcal{L} = \mathcal{H}(t, q^j, \pi_j),$$

of any such a Hamiltonian form H .

It follows that, given Hamiltonian forms H and H' weakly associated with an almost regular Lagrangian L , their difference is $f dt$, ($f \in I_N$). Above proposition enables us to connect Lagrangian and Cartan equations for an almost regular Lagrangian L with the Hamiltonian equations for Hamiltonian forms weakly associated with L [Giachetta *et. al.* (1997)].

Let a section r of $V^*Q \rightarrow \mathbb{R}$ be a solution of the Hamiltonian equations (5.121) for a Hamiltonian form H weakly associated with an almost regular Lagrangian L . If r lives in the constraint space N_L , the section $c = \pi_Q \circ r$ of $Q \rightarrow \mathbb{R}$ satisfies the Lagrangian equations (5.111), while $\bar{c} = \widehat{H} \circ r$ obeys the Cartan equations (5.113).

Given an almost regular Lagrangian L , let a section \bar{c} of the jet bundle $J^1(\mathbb{R}, Q) \rightarrow \mathbb{R}$ be a solution of the Cartan equations (5.113). Let H be a Hamiltonian form weakly associated with L , and let H satisfy the relation

$$\widehat{H} \circ \widehat{L} \circ \bar{c} = j^1(\pi_0^1 \circ \bar{c}). \quad (5.141)$$

Then, the section $r = \widehat{L} \circ \bar{c}$ of the Legendre bundle $V^*Q \rightarrow \mathbb{R}$ is a solution of the Hamiltonian equations (5.121) for H . Since $\widehat{H} \circ \widehat{L}$ is a projection operator, the condition (5.141) implies that the solution \bar{s} of the Cartan equations is actually an integrable section $\bar{c} = \dot{c}$ where c is a solution of the Lagrangian equations.

Given a Hamiltonian form H weakly associated with an almost regular Lagrangian L , let us consider the corresponding constrained Hamiltonian form H_N (5.133). H_N is the same for all Hamiltonian forms weakly associ-

ated with L , and $H_L = \widehat{L}^* H_N$.

For any Hamiltonian form H weakly associated with an almost regular Lagrangian L , every solution of the Hamiltonian equations which lives in the Lagrangian constraint space N_L is a solution of the constrained Hamiltonian equations (5.134).

Using the equality $H_L = \widehat{L}^* H_N$, one can show that the constrained Hamiltonian equations (5.134) are equivalent to the *Hamilton-de Donder equations* (5.115) and are quasi-equivalent to the Cartan equations (5.114) [Giachetta *et. al.* (1997); Mangiarotti and Sardanashvily (1998); Lopez and Marsden (2003)].

5.6.11 Quadratic Degenerate Lagrangian Systems

Given a configuration bundle $Q \rightarrow \mathbb{R}$, let us consider a quadratic Lagrangian L which has the coordinate expression

$$\mathcal{L} = \frac{1}{2} a_{ij} \dot{q}^i \dot{q}^j + b_i \dot{q}^i + c, \quad (5.142)$$

where a , b and c are local functions on Q . This property is coordinate-independent due to the affine transformation law of the coordinates \dot{q}^i . The associated Legendre map

$$p_i \circ \widehat{L} = a_{ij} \dot{q}^j + b_i \quad (5.143)$$

is an affine map over Q . It defines the corresponding linear map

$$\bar{L} : VQ \longrightarrow V^*Q, \quad p_i \circ \bar{L} = a_{ij} \dot{q}^j. \quad (5.144)$$

Let the Lagrangian L (5.142) be almost regular, i.e., the matrix function a_{ij} is of constant rank. Then the Lagrangian constraint space $N_L = \widehat{L}(J^1(\mathbb{R}, Q))$ is an affine subbundle of the bundle $V^*Q \rightarrow Q$, modelled over the vector subbundle \bar{N}_L (5.144) of $V^*Q \rightarrow Q$. Hence, $N_L \rightarrow Q$ has a global section. For the sake of simplicity, let us assume that it is the canonical zero section $\widehat{0}(Q)$ of $V^*Q \rightarrow Q$. Then $\bar{N}_L = N_L$. Therefore, the kernel of the Legendre map (5.143) is an affine subbundle of the affine jet bundle $J^1(\mathbb{R}, Q) \rightarrow Q$, modelled over the kernel of the linear map \bar{L} (5.144). Then there exists a connection Γ on the fibre bundle $Q \rightarrow \mathbb{R}$, given by

$$\Gamma : Q \longrightarrow \text{Ker } \widehat{L} \subset J^1(\mathbb{R}, Q), \quad \text{with} \quad a_{ij} \Gamma_{\mu}^j + b_i = 0.$$

Connections Γ constitute an affine space modelled over the linear space of vertical vector-fields v on $Q \rightarrow \mathbb{R}$, satisfying the conditions

$$a_{ij}v^j = 0 \quad (5.145)$$

and, as a consequence, the conditions $v^i b_i = 0$. If the Lagrangian (5.142) is regular, the connection Γ is unique.

There exists a linear bundle map

$$\sigma : V^*Q \longrightarrow VQ, \quad \dot{q}^i \circ \sigma = \sigma^{ij} p_j,$$

such that $\bar{L} \circ \sigma \circ i_N = i_N$. The map σ is a solution of the algebraic equations

$$a_{ij}\sigma^{jk}a_{kb} = a_{ib}.$$

There exist the bundle splitting

$$VQ = \text{Ker } a \oplus E' \quad (5.146)$$

and a nonholonomic atlas of this bundle such that transition functions of $\text{Ker } a$ and E' are independent. Since a is a non-degenerate fibre metric in E' , there exists an atlas of E' such that a is brought into a diagonal matrix with non-vanishing components a_{AA} . Due to the splitting (5.146), we have the corresponding bundle splitting

$$V^*Q = (\text{Ker } a)^* \oplus \text{Im } a. \quad (5.147)$$

Then the desired map σ is represented by a direct sum $\sigma_1 \oplus \sigma_0$ of an arbitrary section σ_1 of the bundle $\vee^2 \text{Ker } a^* \rightarrow Q$ and the section σ_0 of the bundle $\wedge^2 E' \rightarrow Q$, which has non-vanishing components $\sigma^{AA} = (a_{AA})^{-1}$ with respect to the above atlas of E' . Moreover, σ satisfies the particular relations

$$\sigma_0 = \sigma_0 \circ \bar{L} \circ \sigma_0, \quad a \circ \sigma_1 = 0, \quad \sigma_1 \circ a = 0. \quad (5.148)$$

The splitting (5.146) leads to the splitting

$$J^1(\mathbb{R}, Q) = \mathcal{S}(J^1(\mathbb{R}, Q)) \oplus \mathcal{F}(J^1(\mathbb{R}, Q)) = \text{Ker } \hat{L} \oplus \text{Im}(\sigma \circ \hat{L}) \quad (5.149)$$

$$\dot{q}^i = \mathcal{S}^i + \mathcal{F}^i = [\dot{q}^i - \sigma_0^{ik}(a_{kj}\dot{q}^j + b_k)] + [\sigma_0^{ik}(a_{kj}\dot{q}^j + b_k)], \quad (5.150)$$

while the splitting (5.147) can be written as

$$V^*Q = \mathcal{R}(V^*Q) \oplus \mathcal{P}(V^*Q) = \text{Ker } \sigma_0 \oplus N_L, \quad (5.151)$$

$$p_i = \mathcal{R}_i + \mathcal{P}_i = [p_i - a_{ij}\sigma_0^{jk}p_k] + [a_{ij}\sigma_0^{jk}p_k]. \quad (5.152)$$

Note that, with respect to the coordinates \mathcal{S}_α^i and \mathcal{F}_α^i (5.150), the Lagrangian (5.142) reads

$$\mathcal{L} = \frac{1}{2} a_{ij} \mathcal{F}^i \mathcal{F}^j + c',$$

while the Lagrangian constraint space is given by the *reducible constraints*

$$\mathcal{R}_i = p_i - a_{ij} \sigma_0^{jk} p_k = 0.$$

Given the linear map σ and the connection Γ as defined above, let us consider the affine *Hamiltonian map*

$$\Phi = \widehat{\Gamma} + \sigma : V^*Q \longrightarrow J^1(\mathbb{R}, Q), \quad \Phi^i = \Gamma^i + \sigma^{ij} p_j, \quad (5.153)$$

and the *Hamiltonian form*

$$\begin{aligned} H = H_\Phi + \Phi^* L &= p_i dq^i - [p_i \Gamma^i + \frac{1}{2} \sigma_0^{ij} p_i p_j + \sigma_1^{ij} p_i p_j - c'] dt \\ &= (\mathcal{R}_i + \mathcal{P}_i) dq^i - [(\mathcal{R}_i + \mathcal{P}_i) \Gamma^i + \frac{1}{2} \sigma_0^{ij} \mathcal{P}_i \mathcal{P}_j + \sigma_1^{ij} p_i p_j - c'] dt. \end{aligned} \quad (5.154)$$

In particular, if σ_1 is non-degenerate, so is the Hamiltonian form H .

The Hamiltonian forms of the type H , parameterized by connections Γ , are weakly associated with the Lagrangian (5.142) and constitute a complete set. Then H is weakly associated with L . Let us write the corresponding Hamiltonian equations (5.121) for a section r of the Legendre bundle $V^*Q \rightarrow \mathbb{R}$. They are

$$\dot{c} = (\widehat{\Gamma} + \sigma) \circ r, \quad c = \pi_Q \circ r. \quad (5.155)$$

Due to the surjections \mathcal{S} and \mathcal{F} (5.150), the Hamiltonian equations (5.155) break in two parts

$$\mathcal{S} \circ \dot{c} = \Gamma \circ c, \quad \dot{r}^i - \sigma^{ik} (a_{kj} \dot{r}^j + b_k) = \Gamma^i \circ c, \quad (5.156)$$

$$\mathcal{F} \circ \dot{c} = \sigma \circ r, \quad \sigma^{ik} (a_{kj} \dot{r}^j + b_k) = \sigma^{ik} r_k. \quad (5.157)$$

Let c be an arbitrary section of $Q \rightarrow \mathbb{R}$, e.g., a solution of the Lagrangian equations. There exists a connection Γ such that the relation (5.156) holds, namely, $\Gamma = \mathcal{S} \circ \Gamma'$, where Γ' is a connection on $Q \rightarrow \mathbb{R}$ which has c as an integral section.

If $\sigma_1 = 0$, then $\Phi = \widehat{H}$ and the Hamiltonian forms H are associated with the Lagrangian (5.142). Thus, for different σ_1 , we have different complete sets of Hamiltonian forms H , which differ from each other in the term $v^i \mathcal{R}_i$, where v are vertical vector-fields (5.145). This term vanishes on the

Lagrangian constraint space. The corresponding constrained Hamiltonian form $H_N = i_N^* H$ and the constrained Hamiltonian equations (5.134) can be written.

For every Hamiltonian form H , the Hamiltonian equations (5.121) and (5.157) restricted to the Lagrangian constraint space N_L are equivalent to the constrained Hamiltonian equations.

Due to the splitting (5.151), we have the corresponding splitting of the vertical tangent bundle $V_Q V^* Q$ of the bundle $V^* Q \rightarrow Q$. In particular, any vertical vector-field u on $V^* Q \rightarrow \mathbb{R}$ admits the decomposition

$$u = [u - u_{TN}] + u_{TN}, \quad \text{with} \quad u_{TN} = u^i \partial_i + a_{ij} \sigma_0^{jk} u_k \partial^i,$$

such that $u_N = u_{TN}|_{N_L}$ is a vertical vector-field on the Lagrangian constraint space $N_L \rightarrow \mathbb{R}$. Let us consider the equations

$$r^*(u_{TN} \rfloor dH) = 0$$

where r is a section of $V^* Q \rightarrow \mathbb{R}$ and u is an arbitrary vertical vector-field on $V^* Q \rightarrow \mathbb{R}$. They are equivalent to the pair of equations

$$r^*(a_{ij} \sigma_0^{jk} \partial^i \rfloor dH) = 0, \quad (5.158)$$

$$r^*(\partial_i \rfloor dH) = 0. \quad (5.159)$$

Restricted to the Lagrangian constraint space, the Hamiltonian equations for different Hamiltonian forms H associated with the same quadratic Lagrangian (5.142) differ from each other in the equations (5.156). These equations are independent of momenta and play the role of *gauge-type conditions*.

5.6.12 Time-Dependent Integrable Hamiltonian Systems

Recall that the configuration space of a time-dependent mechanical system is a fibre bundle $M \rightarrow \mathbb{R}$ over the time axis \mathbb{R} equipped with the bundle coordinates $q^\alpha \equiv (t, q^k)$, for $k = 1, \dots, m$. The corresponding momentum phase-space is the vertical cotangent bundle $V^* M$ of $M \rightarrow \mathbb{R}$ with holonomic bundle coordinates (t, q^k, p_k) .

Recall that the cotangent bundle $T^* M$ of M is coordinated by [Mangiarotti and Sardanasvily (1998); Giachetta *et. al.* (1997)]

$$(t, q^k, p_0 = p, p_k), \quad p' = p + \frac{\partial q^k}{\partial t} p_k, \quad (5.160)$$

and plays the role of the homogeneous momentum phase-space of time-dependent mechanics. It admits the canonical Liouville form $\Xi = p_\alpha dq^\alpha$, the canonical symplectic form $\Omega_T = d\Xi$, and the corresponding Poisson bracket

$$\{f, f'\}_T = \partial^\alpha f \partial_\alpha f' - \partial_\alpha f \partial^\alpha f', \quad (f, f' \in C^\infty(T^*M)). \quad (5.161)$$

There is a canonical 1D fibre bundle

$$\zeta : T^*M \rightarrow V^*M, \quad (5.162)$$

whose kernel is the annihilator of the vertical tangent bundle $VM \subset TM$. The transformation law (5.160) shows that it is a trivial affine bundle. Indeed, given a global section h of ζ , one can equip T^*M with the fibre coordinate $r = p - h$ possessing the identity transition functions.

The fibre bundle (5.162) gives the vertical cotangent bundle V^*M with the canonical Poisson structure $\{\cdot, \cdot\}_V$ such that

$$\zeta^* \{f, f'\}_V = \{\zeta^* f, \zeta^* f'\}_T, \quad (5.163)$$

$$\{f, f'\}_V = \partial^k f \partial_k f' - \partial_k f \partial^k f', \quad (5.164)$$

for all $f, f' \in C^\infty(V^*M)$. The corresponding symplectic foliation coincides with the fibration $V^*M \rightarrow \mathbb{R}$.

However, the Poisson structure (5.164) fails to give any dynamical equation on the momentum phase-space V^*M because Hamiltonian vector-fields

$$\vartheta_f = \partial^k f \partial_k - \partial_k f \partial^k, \quad \vartheta_f \rfloor df' = \{f, f'\}_V, \quad (f, f' \in C^\infty(V^*M)),$$

of functions on V^*M are vertical. Hamiltonian dynamics of time-dependent mechanics is described in a different way as a particular Hamiltonian dynamics on fibre bundles [Mangiarotti and Sardanašvily (1998); Giachetta *et. al.* (1997)].

A Hamiltonian on the momentum phase-space $V^*M \rightarrow \mathbb{R}$ of time-dependent mechanics is defined as a global section

$$h : V^*M \rightarrow T^*M, \quad p \circ h = -\mathcal{H}(t, q^j, p_j),$$

of the affine bundle ζ (5.162). It induces the pull-back Hamiltonian form

$$H = h^* \Xi = p_k dq^k - \mathcal{H} dt, \quad (5.165)$$

on V^*M . Given H (5.165), there exists a unique vector-field γ_H on V^*M such that

$$\gamma_H \rfloor dt = 1, \quad \gamma_H \rfloor dH = 0. \quad (5.166)$$

This vector-field reads

$$\gamma_H = \partial_t + \partial^k \mathcal{H} \partial_k - \partial_k \mathcal{H} \partial^k. \quad (5.167)$$

It defines the first-order Hamiltonian equation

$$\dot{t} = 1, \quad \dot{q}^k = \partial^k \mathcal{H}, \quad \dot{p}_k = -\partial_k \mathcal{H} \quad (5.168)$$

on V^*M , where $(t, q^k, p_k, \dot{t}, \dot{q}^k, \dot{p}_k)$ are holonomic coordinates on the tangent bundle TV^*M . Solutions of this equation are trajectories of the vector-field γ_H . They assemble into a (regular) foliation of V^*M .

A first integral of the Hamiltonian equation (5.168) is defined as a smooth real function F on V^*M whose Lie derivative

$$\mathcal{L}_{\gamma_H} F = \gamma_H \rfloor dF = \partial_t F + \{\mathcal{H}, F\}_V$$

along the vector-field γ_H (5.167) vanishes, i.e., the function F is constant on trajectories of the vector-field γ_H . A time-dependent Hamiltonian system (V^*M, H) on V^*M is said to be completely integrable if the Hamiltonian equation (5.168) admits m first integrals F_k which are in involution with respect to the Poisson bracket $\{\cdot, \cdot\}_V$ (5.164) and whose differentials dF_k are linearly independent almost everywhere. This system can be extended to an autonomous completely integrable Hamiltonian system on T^*M as follows.

Let us consider the pull-back ζ^*H of the Hamiltonian form $H = h^*\Xi$ onto the cotangent bundle T^*M . Note that the difference $\Xi - \zeta^*h^*\Xi$ is a horizontal 1-form on $T^*M \rightarrow \mathbb{R}$ and that

$$\mathcal{H}^* = \partial_t \rfloor (\Xi - \zeta^*h^*\Xi) = p + \mathcal{H} \quad (5.169)$$

is a function on T^*M [Sniatycki (1980)]. Let us regard \mathcal{H}^* (5.169) as a Hamiltonian of an autonomous Hamiltonian system on the symplectic manifold (T^*M, Ω_T) . The Hamiltonian vector-field of \mathcal{H}^* on T^*M reads

$$\gamma_T = \partial_t - \partial_t \mathcal{H} \partial^0 + \partial^k \mathcal{H} \partial_k - \partial_k \mathcal{H} \partial^k.$$

It is projected onto the vector-field γ_H (5.167) on V^*M , and the relation

$$\zeta^*(\mathcal{L}_{\gamma_H} f) = \{\mathcal{H}^*, \zeta^* f\}_T, \quad (f \in C^\infty(V^*M)).$$

holds. An immediate consequence of this relation is the following.

Let $(V^*M, H; F_k)$ be a time-dependent completely integrable Hamiltonian system with first integrals $\{F_k\}$ on V^*M . Then $(T^*M; \mathcal{H}^*, \zeta^*F_k)$ is an autonomous completely integrable Hamiltonian system on T^*M whose first integrals $\{\mathcal{H}^*, \zeta^*F_k\}$ are in involution with respect to the Poisson bracket $\{, \}_T$ (5.161). Furthermore, let N be a connected invariant manifold of the time-dependent completely integrable Hamiltonian system $(V^*M, H; F_k)$. Then $h(N) \subset T^*M$ is a connected invariant manifold of the completely integrable Hamiltonian system $(T^*M; \mathcal{H}^*, \zeta^*F_k)$ on T^*M . If N contains no critical points of first integrals F_k , then $\{\mathcal{H}^*, \zeta^*F_k\}$ have no critical points in $h(N)$.

5.6.13 Time-Dependent Action-Angle Coordinates

Let us introduce *time-dependent action-angle coordinates* around an invariant manifold N of a time-dependent completely integrable Hamiltonian system (V^*M, H) as those induced by the action-angle coordinates around the invariant manifold $h(N)$ of the autonomous completely integrable system (T^*M, \mathcal{H}^*) .

Let M' be a connected invariant manifold of an autonomous completely integrable system (F_α) , $\alpha = 1, \dots, n$, on a symplectic manifold (Z, Ω_Z) , and let the Hamiltonian vector-fields of first integrals F_α on M' be complete. Let there exist a neighborhood U of M' such that F_α have no critical points in U and the submersion $\times F_\alpha : U \rightarrow \mathbb{R}^n$ is a trivial fibre bundle over a domain $V' \subset \mathbb{R}^n$. Then U is isomorphic to the *symplectic annulus*

$$W = V' \times (\mathbb{R}^{n-m} \times T^m),$$

provided with the generalized action-angle coordinates

$$(I_1, \dots, I_n; x^1, \dots, x^{n-m}; \phi^1, \dots, \phi^m)$$

such that the symplectic form on W reads

$$\Omega_Z = dI_i \wedge x^i + dI_{n-m+k} \wedge d\phi^k,$$

and the first integrals F_α are functions of the action coordinates (I_α) only.

In particular, let M' be a compact invariant manifold of a completely integrable system $\{F_\alpha\}$, $\alpha = 1, \dots, n$, on a symplectic manifold (Z, Ω_Z) which does not contain critical points of the first integrals F_α . Let the vector-field γ_H (5.167) be complete. Let a connected invariant manifold N of a time-dependent completely integrable Hamiltonian system $(V^*M, H; F_k)$

contain no critical points of first integrals F_k , and let its projection N_0 onto the fibre V_0^*M along trajectories of γ_H be compact. Then the invariant manifold $h(N)$ of the completely integrable Hamiltonian system $(T^*M; \mathcal{H}^*, \zeta^*F_k)$ has an open neighborhood U .

Now, the open neighborhood U of the invariant manifold $h(N)$ of the completely integrable Hamiltonian system $(T^*M; \mathcal{H}^*, \zeta^*F_k)$ is isomorphic to the symplectic annulus

$$W' = V' \times (\mathbb{R} \times T^m), \quad V' = (-\varepsilon, \varepsilon) \times V, \quad (5.170)$$

provided with the generalized action–angle coordinates

$$(I_0, \dots, I_m; t, \phi^1, \dots, \phi^m). \quad (5.171)$$

Moreover, we find that $J_0 = r$, $a_0^\alpha = \delta_0^\alpha$ and, as a consequence,

$$a_0^0 = \frac{\partial I_0}{\partial J_0} = 1, \quad a_i^0 = \frac{\partial I_i}{\partial J_0} = 0,$$

i.e., the action coordinate I_0 is linear in the coordinate r , while I_i are independent of r . With respect to the coordinates (5.171), the symplectic form on W' reads

$$\Omega_T = dI_0 \wedge dt + dI_k \wedge d\phi^k,$$

the Hamiltonian \mathcal{H}^* is an affine function $\mathcal{H}^* = I_0 + \mathcal{H}'(I_j)$ of the action coordinate I_0 , while the first integrals ζ^*F_k depends only on the action coordinates I_i . The Hamiltonian vector–field of the Hamiltonian \mathcal{H}^* is

$$\gamma_T = \partial_t + \partial^i \mathcal{H}' \partial_i. \quad (5.172)$$

Since the action coordinates I_i are independent on the coordinate r , the symplectic annulus W' (5.170) inherits the composite fibration

$$W' \rightarrow V \times (\mathbb{R} \times T^m) \rightarrow \mathbb{R}. \quad (5.173)$$

Therefore, one can regard $W = V \times (\mathbb{R} \times T^m)$ as a momentum phase–space of the time–dependent Hamiltonian system in question around the invariant manifold N . It is coordinated by (I_i, t, ϕ^i) , which we agree to call the time–dependent action–angle coordinates. By the relation similar to (5.163), W can be equipped with the Poisson structure

$$\{f, f'\}_W = \partial^i f \partial_i f' - \partial_i f \partial^i f',$$

while the global section $h' : W \rightarrow W'$ such that $I_0 \circ h' = -\mathcal{H}'$, of the trivial bundle ζ (5.173), gives W with the Hamiltonian form

$$H' = I_i d\phi^i - \mathcal{H}'(I_j) dt.$$

The associated vector-field γ_H (5.166) is exactly the projection onto W of the Hamiltonian vector-field γ_T (5.172), and takes the same coordinate form. It defines the Hamiltonian equation on W ,

$$I_i = \text{const}, \quad \dot{\phi}^i = \partial^i \mathcal{H}'(I_j).$$

One can think of this equation as being the Hamiltonian equation of a time-dependent Hamiltonian system around the invariant manifold N relative to time-dependent action-angle coordinates.

5.6.14 *Lyapunov Stability*

The notion of the Lyapunov stability of a dynamical equation on a smooth manifold implies that this manifold is equipped with a Riemannian metric. At the same time, no preferable Riemannian metric is associated to a first-order dynamical equation. Here, we aim to study the Lyapunov stability of first-order dynamical equations in non-autonomous mechanics with respect to different (time-dependent) Riemannian metrics.

Let us recall that a solution $s(t)$, for all $t \in \mathbb{R}$, of a first-order dynamical equation is said to be Lyapunov stable (in the positive direction) if for $t_0 \in \mathbb{R}$ and any $\varepsilon > 0$, there is $\delta > 0$ such that, if $s'(t)$ is another solution and the distance between the points $s(t_0)$ and $s'(t_0)$ is inferior to δ , then the distance between the points $s(t)$ and $s'(t)$ for all $t > t_0$ is inferior to ε . In order to formulate a criterion of the Lyapunov stability with respect to a time-dependent Riemannian metric, we introduce the notion of a covariant Lyapunov tensor as generalization of the well-known Lyapunov matrix. The latter is defined as the coefficient matrix of the variation equation [Gallavotti (1983); Hirsch and Smale (1974)], and fails to be a tensor under coordinate transformations, unless they are linear and time-independent. On the contrary, the covariant Lyapunov tensor is a true tensor-field, but it essentially depends on the choice of a Riemannian metric. The following was shown in [Sardanashvily (2002b)]:

(i) If the covariant Lyapunov tensor is negative definite in a tubular neighborhood of a solution s at points $t \geq t_0$, this solution is Lyapunov stable.

(ii) For any first-order dynamical equation, there exists a (time-dependent) Riemannian metric such that every solution of this equation is Lyapunov stable.

(iii) Moreover, the Lyapunov exponent of any solution of a first-order dynamical equation can be made equal to any real number with respect to the appropriate (time-dependent) Riemannian metric. It follows that chaos in dynamical systems described by smooth (C^∞) first-order dynamical equations can be characterized in full by time-dependent Riemannian metrics.

5.6.15 First-Order Dynamical Equations

Let \mathbb{R} be the time axis provided with the Cartesian coordinate t and transition functions $t' = t + \text{const}$. In geometrical terms [Mangiarotti and Sardanashevily (1998)], a (smooth) first-order dynamical equation in non-autonomous mechanics is defined as a vector-field γ on a smooth fibre bundle

$$\pi : Y \rightarrow \mathbb{R} \quad (5.174)$$

which obeys the condition $\gamma \rfloor dt = 1$, i.e.,

$$\gamma = \partial_t + \gamma^k \partial_k. \quad (5.175)$$

The associated first-order dynamical equation takes the form

$$\dot{t} = 1, \quad \dot{y}^k = \gamma^k(t, y^j) \partial_k, \quad (5.176)$$

where $(t, y^k, \dot{t}, \dot{y}^k)$ are holonomic coordinates on TY . Its solutions are trajectories of the vector-field γ (5.175). They assemble into a (regular) foliation \mathcal{F} of Y . Equivalently, γ (5.175) is defined as a connection on the fibre bundle (5.174).

A fibre bundle Y (5.174) is trivial, but it admits different trivializations

$$Y \cong \mathbb{R} \times M, \quad (5.177)$$

distinguished by fibrations $Y \rightarrow M$. For example, if there is a trivialization (5.177) such that, with respect to the associated coordinates, the components γ^k of the connection γ (5.175) are independent of t , one says that γ is a conservative first-order dynamical equation on M .

Hereafter, the vector-field γ (5.175) is assumed to be *complete*, i.e., there is a unique global solution of the dynamical equation γ through each

point of Y . For example, if fibres of $Y \rightarrow \mathbb{R}$ are compact, any vector-field γ (5.175) on Y is complete.

If the vector-field γ (5.175) is complete, there exists a trivialization (5.177) of Y , with an atlas $\Psi = \{(U; t, y^a)\}$ of a fibre bundle $Y \rightarrow \mathbb{R}$ with time-independent transition functions $y'^a(y^b)$, such that any solution s of γ reads

$$s^a(t) = \text{const}, \quad (\text{for all } t \in \mathbb{R}),$$

with respect to associated bundle coordinates (t, y^a) . If γ is complete, the foliation \mathcal{F} of its trajectories is a fibration ζ of Y along these trajectories onto any fibre of Y , e.g., $Y_{t=0} \cong M$. This fibration induces a desired trivialization [Mangiarotti and Sardanashvily (1998)].

One can think of the coordinates (t, y^a) as being the *initial-date coordinates* because all points of the same trajectory differ from each other only in the temporal coordinate.

Let us consider the canonical lift $V\gamma$ of the vector-field γ (5.175) onto the vertical tangent bundle VY of $Y \rightarrow \mathbb{R}$. With respect to the holonomic bundle coordinates (t, y^k, \bar{y}^k) on VY , it reads

$$V\gamma = \gamma + \partial_j \gamma^k \bar{y}^j \bar{\partial}_k, \quad \text{where} \quad \bar{\partial}_k = \frac{\partial}{\partial \bar{y}^k}.$$

This vector-field obeys the condition $V\gamma|dt = 1$, and defines the first-order dynamical equation

$$\dot{t} = 1, \quad \dot{y}^k = \gamma^k(t, y^i), \quad (5.178)$$

$$\dot{\bar{y}}^k = \partial_j \gamma^k(t, y^i) \bar{y}^j \quad (5.179)$$

on VY . The equation (5.178) coincides with the initial one (5.176). The equation (5.179) is the well-known variation equation. Substituting a solution s of the initial dynamical equation (5.178) into (5.179), one gets a linear dynamical equation whose solutions \bar{s} are Jacobi fields of the solution s . In particular, if $Y \rightarrow \mathbb{R}$ is a vector bundle, there are the canonical splitting $VY \cong Y \times Y$ and the map $VY \rightarrow Y$ so that $s + \bar{s}$ obeys the initial dynamical equation (5.178) modulo the terms of order > 1 in \bar{s} .

5.6.16 *Lyapunov Tensor and Stability*

5.6.16.1 *Lyapunov Tensor*

The collection of coefficients

$$l_j^k = \partial_j \gamma^k \quad (5.180)$$

of the variation equation (5.179) is called the *Lyapunov matrix*. Clearly, it is not a tensor under bundle coordinate transformations of the fibre bundle Y (5.174). Therefore, we introduce a *covariant Lyapunov tensor* as follows.

Let a fibre bundle $Y \rightarrow \mathbb{R}$ be provided with a Riemannian fibre metric g , defined as a section of the *symmetrized tensor product*

$$\vee^2 V^*Y \rightarrow Y \quad (5.181)$$

of the vertical cotangent bundle V^*Y of $Y \rightarrow \mathbb{R}$. With respect to the holonomic coordinates (t, y^k, \bar{y}_k) on V^*Y , it takes the coordinate form

$$g = \frac{1}{2} g_{ij}(t, y^k) \bar{dy}^i \vee \bar{dy}^j,$$

where $\{\bar{dy}^i\}$ are the holonomic fibre bases for V^*Y .

Given a first-order dynamical equation γ , let

$$V^*\gamma = \gamma - \partial_j \gamma^k \bar{y}_k \bar{\partial}^j, \quad \text{where} \quad \bar{\partial}^j = \frac{\partial}{\partial \bar{y}_j}. \quad (5.182)$$

be the canonical lift of the vector-field γ (5.175) onto V^*Y . It is a connection on $V^*Y \rightarrow \mathbb{R}$. Let us consider the Lie derivative $\mathfrak{L}_\gamma g$ of the Riemannian fibre metric g along the vector-field $V^*\gamma$ (5.182). It reads

$$L_{ij} = (\mathfrak{L}_\gamma g)_{ij} = \partial_t g_{ij} + \gamma^k \partial_k g_{ij} + \partial_i \gamma^k g_{kj} + \partial_j \gamma^k g_{ik}. \quad (5.183)$$

This is a section of the fibre bundle (5.181) and, consequently, a tensor with respect to any bundle coordinate transformation of the fibre bundle (5.174). We agree to call it the covariant Lyapunov tensor. If g is an Euclidean metric, it becomes the following symmetrization of the Lyapunov matrix (5.180),

$$L_{ij} = \partial_i \gamma^j + \partial_j \gamma^i = l_i^j + l_j^i.$$

Let us point the following two properties of the covariant Lyapunov tensor.

(i) Written with respect to the initial–date coordinates, the covariant Lyapunov tensor is given by

$$L_{ab} = \partial_t g_{ab}.$$

(ii) Given a solution s of the dynamical equation γ and a solution \bar{s} of the variation equation (5.179), we have

$$L_{ij}(t, s^k(t)) \bar{s}^i \bar{s}^j = \partial_t (g_{ij}(t, s^k(t)) \bar{s}^i \bar{s}^j).$$

The definition of the covariant Lyapunov tensor (5.183) depends on the choice of a Riemannian fibre metric on the fibre bundle Y .

If the vector–field γ is complete, there is a Riemannian fibre metric on Y such that the covariant Lyapunov tensor vanishes everywhere. Let us choose the atlas of the initial–date coordinates. Using the fibration $\zeta : Y \rightarrow Y_{t=0}$, one can give Y with a time–independent Riemannian fibre metric

$$g_{ab}(t, y^c) = h(t) g_{ab}(0, y^c) \quad (5.184)$$

where $g_{ab}(0, y^c)$ is a Riemannian metric on the fibre $Y_{t=0}$ and $h(t)$ is a positive smooth function on \mathbb{R} . The covariant Lyapunov tensor with respect to the metric (5.184) is given by

$$L_{ab} = \partial_t h g_{ab}.$$

Putting $h(t) = 1$, we get $L = 0$.

5.6.16.2 Lyapunov Stability

With the covariant Lyapunov tensor (5.183), we get the following criterion of the stability condition of Lyapunov.

Recall that, given a Riemannian fibre metric g on a fibre bundle $Y \rightarrow \mathbb{R}$, the instant–wise distance $\rho_t(s, s')$ between two solutions s and s' of a dynamical equation γ on Y at an instant t is the distance between the points $s(t)$ and $s'(t)$ in the Riemannian space $(Y_t, g(t))$.

Let s be a solution of a first–order dynamical equation γ . If there exists an open tubular neighborhood U of the trajectory s where the covariant Lyapunov tensor (5.183) is negative-definite at all instants $t \geq t_0$, then there exists an open tubular neighborhood U' of s such that

$$\lim_{t' \rightarrow \infty} [\rho_{t'}(s, s') - \rho_t(s, s')] < 0$$

for any $t > t_0$ and any solution s' crossing U' . Since the condition and the statement are coordinate-independent, let us choose the following chart of initial-date coordinates that covering the trajectory s . Put $t = 0$ without a loss of generality. There is an open neighborhood $U_0 \subset Y_0 \cap U$ of $s(0)$ in the Riemannian manifold $(Y_0, g(0))$ which can be provided with the normal coordinates (x^a) defined by the Riemannian metric $g(0)$ in Y_0 and centralized at $s(0)$. Let us consider the open tubular $U' = \zeta^{-1}(U_0)$ with the coordinates (t, x^a) . It is the desired chart of initial-date coordinates. With respect to these coordinates, the solution s reads $s^a(t) = 0$. Let $s'^a(t) = u^a = \text{const}$ be another solution crossing U' . The instant-wise distance $\rho_t(s, s')$, $t \geq 0$, between solutions s and s' is the distance between the points $(t, 0)$ and (t, u) in the Riemannian manifold $(Y_t, g(t))$. This distance does not exceed the length

$$\bar{\rho}_t(s, s') = \left[\int_0^1 g_{ab}(t, \tau u^c) u^a u^b d\tau \right]^{1/2} \quad (5.185)$$

of the curve

$$x^a = \tau u^a, \quad (\tau \in [0, 1]) \quad (5.186)$$

in the Riemannian space $(Y_t, g(t))$, while $\rho_0(s, s') = \bar{\rho}_0(s, s')$. The temporal derivative of the function $\bar{\rho}_t(s, s')$ (5.185) reads

$$\partial_t \bar{\rho}_t(s, s') = \frac{1}{2(\bar{\rho}_t(s, s'))^{1/2}} \int_0^1 \partial_t g_{ab}(t, \tau u^c) u^a u^b d\tau. \quad (5.187)$$

Since the bilinear form $\partial_t g_{ab} = L_{ab}$, $t \geq 0$, is negative-definite at all points of the curve (5.186), the derivative (5.187) at all points $t \geq t_0$ is also negative. Hence, we get

$$\rho_{t>0}(s, s') < \bar{\rho}_{t>0}(s, s') < \bar{\rho}_0(s, s') = \rho_0(s, s').$$

The solution s is *Lyapunov stable* with respect to the Riemannian fibre metric g . One can think of the solution s as being *isometrically Lyapunov stable*. Being Lyapunov stable with respect a Riemannian fibre metric g , a solution s need not be so with respect to another Riemannian fibre metric g' , unless g' results from g by a time-independent transformation.

For any first-order dynamical equation defined by a complete vector-field γ (5.175) on a fibre bundle $Y \rightarrow \mathbb{R}$, there exists a Riemannian fibre metric on Y such that each solution of γ is Lyapunov stable. This property

obviously holds with respect to the Riemannian fibre metric (5.184) where $h = 1$.

Let λ be a real number. Given a dynamical equation γ defined by a complete vector-field γ (5.175), there is a Riemannian fibre metric on Y such that the Lyapunov spectrum of any solution of γ reduces to λ . To prove this, recall that the (upper) Lyapunov exponent of a solution s' with respect to a solution s is defined as the limit

$$K(s, s') = \lim_{t \rightarrow \infty} \frac{1}{t} \ln(\rho_t(s, s')). \quad (5.188)$$

Let us give Y with the Riemannian fibre metric (5.184) where $h = \exp(\lambda t)$. A simple computation shows that the Lyapunov exponent (5.188) with respect to this metric is exactly λ .

If the upper limit

$$\lambda = \lim_{\rho_{t=0}(s, s') \rightarrow 0} K(s, s')$$

is *negative*, the solution s is said to be exponentially Lyapunov stable. If there exists at least one positive Lyapunov exponent, one speaks about chaos in a dynamical system [Gutzwiller (1990)]. This shows that chaos in smooth dynamical systems can be characterized in full by time-dependent Riemannian metrics.

Example

Here is a simple example which shows that solutions of a smooth first-order dynamical equation can be made *Lyapunov stable* at will by the choice of an appropriate time-dependent Riemannian metric.

Let \mathbb{R} be the time axis provided with the Cartesian coordinate t . In geometrical terms, a (smooth) first-order dynamical equation in non-autonomous mechanics is defined as a vector-field γ on a smooth fibre bundle $Y \rightarrow \mathbb{R}$ which obeys the condition $\gamma \rfloor dt = 1$. With respect to bundle coordinates (t, y^k) on Y , this vector-field becomes (5.175). The associated first-order dynamical equation takes the form

$$\dot{y}^k = \gamma^k(t, y^j) \partial_k,$$

where $(t, y^k, \dot{t}, \dot{y}^k)$ are holonomic coordinates on the tangent bundle TY of Y . Its solutions are trajectories of the vector-field γ (5.175).

Let a fibre bundle $Y \rightarrow \mathbb{R}$ be provided with a Riemannian fibre metric g , defined as a section of the symmetrized tensor product $\vee^2 V^*Y \rightarrow Y$ of the

vertical cotangent bundle V^*Y of $Y \rightarrow \mathbb{R}$. With respect to the holonomic coordinates (t, y^k, \bar{y}_k) on V^*Y , it takes the coordinate form

$$g = \frac{1}{2} g_{ij}(t, y^k) \bar{d}y^i \vee \bar{d}y^j,$$

where $\{\bar{d}y^i\}$ are the holonomic fibre bases for V^*Y .

Recall that above we have proposed the following: Let λ be a real number. Given a dynamical equation defined by a complete vector-field γ (5.175), there exists a Riemannian fibre metric on Y such that the Lyapunov spectrum of any solution of γ is λ . The following example aims to illustrate this fact.

Let us consider 1D motion on the axis \mathbb{R} defined by the first-order dynamical equation

$$\dot{y} = y \quad (5.189)$$

on the fibre bundle $Y = \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ coordinated by (t, y) . Solutions of the equation (5.189) read

$$s(t) = c \exp(t), \quad (\text{with } c = \text{const}). \quad (5.190)$$

Let $e_{yy} = 1$ be the standard Euclidean metric on \mathbb{R} . With respect to this metric, the instant-wise distance between two arbitrary solutions

$$s(t) = c \exp(t), \quad s'(t) = c' \exp(t) \quad (5.191)$$

of the equation (5.189) is

$$\rho_t(s, s')_e = |c - c'| \exp(t).$$

Hence, the Lyapunov exponent $K(s, s')$ (5.188) equals 1, and so is the Lyapunov spectrum of any solution (5.190) of the first-order dynamical equation (5.189).

Let now λ be an arbitrary real number. There exists a coordinate $y' = y \exp(-t)$ on \mathbb{R} such that, written relative to this coordinate, the solutions (5.190) of the equation (5.189) read $s(t) = \text{const}$. Let us choose the Riemannian fibre metric on $Y \rightarrow \mathbb{R}$ which takes the form $g_{y'y'} = \exp(2\lambda t)$ with respect to the coordinate y' . Then relative to the coordinate y , it reads

$$g_{yy} = \frac{\partial y'}{\partial y} \frac{\partial y'}{\partial y} g_{y'y'} = \exp(2(\lambda - 1)t). \quad (5.192)$$

The instant-wise distance between the solutions s and s' (5.191) with respect to the metric g (5.192) is

$$\rho_t(s, s') = [g_{yy}(s(t) - s'(t))^2]^{1/2} = |c - c'| \exp(\lambda t).$$

One at once gets that the Lyapunov spectrum of any solution of the differential equation (5.189) with respect to the metric (5.192) is λ .

5.7 Application: Jets and Multi-Time Rheonomic Dynamics

Recall that a number of geometrical models in mechanics and physics are based on the notion of ordinary, autonomous Lagrangian (i.e., a smooth real function on $\mathbb{R} \times TM$). In this sense, we recall that a *Lagrangian space* $L^n = (M, L(x, y))$ is defined as a pair which consists of a real, smooth, n D manifold M with local coordinates $x^i, (i = 1, \dots, n)$ and a regular Lagrangian $L : TM \rightarrow \mathbb{R}$. The geometry of Lagrangian spaces is now used in various fields to study natural phenomena where the dependence on position, velocity or momentum is involved [Kamron and Olver (1989)]. Also, this geometry gives a model for both the gravitational and electromagnetic field theory, in a very natural blending of the geometrical structure of the space with the characteristic properties of the physical fields. Again, there are many problems in physics and variational calculus in which time-dependent Lagrangians are involved.

In the context exposed in [Miron *et. al.* (1988); Miron and Anastasiei (1994)], the energy action functional E , attached to a given time-dependent Lagrangian,

$$L : \mathbb{R} \times TM \rightarrow \mathbb{R}, \quad (t, x^i, v^i) \mapsto L(t, x^i, v^i), \quad (i = 1, \dots, n)$$

not necessarily homogenous with respect to the direction $\{v^i\}$, is of the form

$$E(c) = \int_a^b L(t, x^i(t), \dot{x}^i(t)) dt, \quad (5.193)$$

where $[a, b] \subset \mathbb{R}$, and $c : [a, b] \rightarrow M$ is a smooth curve, locally expressed by $t \mapsto x^i(t)$, and having the velocity $\dot{x} = (\dot{x}^i(t))$. It is obvious that the non-homogeneity of the Lagrangian L , regarded as a smooth function on the product manifold $\mathbb{R} \times TM$, implies that the energy action functional E is dependent of the parametrizations of every curve c . In order to remove this

difficulty, [Miron *et. al.* (1988); Miron and Anastasiei (1994)] regard the space $\mathbb{R} \times TM$ like a fibre bundle over M . In this context, the *geometrical invariance group* of $\mathbb{R} \times TM$ is given by

$$\bar{t} = t, \quad \bar{x}^i = x^i(x^j), \quad \bar{v}^i = \frac{\partial \bar{x}^i}{\partial x^j} v^j. \quad (5.194)$$

The structure of the gauge group (5.194) emphasizes the absolute character of the time t from the classical rheonomic Lagrangian mechanics. At the same time, we point out that the gauge group (5.194) is a subgroup of the *gauge group* of the configuration bundle $J^1(\mathbb{R}, M)$, given as

$$\bar{t} = \bar{t}(t), \quad \bar{x}^i = x^i(x^j), \quad \bar{v}^i = \frac{\partial \bar{x}^i}{\partial x^j} \frac{dt}{d\bar{t}} v^j. \quad (5.195)$$

In other words, the gauge group (5.195) of the jet bundle $J^1(\mathbb{R}, M)$, from the relativistic rheonomic Lagrangian mechanics is more general than that used in the classical rheonomic Lagrangian mechanics, which ignores the temporal reparametrizations. A deep exposition of the physical aspects of the classical rheonomic Lagrangian geometry is done by [Ikeda (1990)], while the classical rheonomic Lagrangian mechanics is done by [Matsumoto (1982)].

Therefore, to remove the parametrization dependence of E , they ignore the temporal repametrizations on $\mathbb{R} \times TM$. Naturally, in these conditions, their energy functional becomes a well defined one, but their approach stands out by the ‘absolute’ character of the time t .

In a more general geometrical approach, [Neagu and Udriște (2000a); Udriște (2000); Neagu (2002)] tried to remove this inconvenience. Following this approach, we regard the mechanical 1-jet space $J^1(\mathbb{R}, M) \equiv \mathbb{R} \times TM$ as a fibre bundle over the base product-manifold $\mathbb{R} \times M$. The gauge group of this bundle of configurations is given by 5.195. Consequently, our gauge group does not ignore the temporal reparametrizations, hence, it stands out by the relativistic character of the time t . In these conditions, using a given semi-Riemannian metric $h_{11}(t)$ on \mathbb{R} , we construct the more general and natural energy action functional, setting

$$E(c) = \int_a^b L(t, x^i(t), \dot{x}^i(t)) \sqrt{|h_{11}|} dt. \quad (5.196)$$

Obviously, E is well defined and is independent of the curve parametrizations.

5.7.1 Relativistic Rheonomic Lagrangian Spaces

In order to develop the time-dependent Lagrangian geometry, following [Neagu and Udriște (2000a); Udriște (2000); Neagu (2002); Neagu and Udriște (2000b); Neagu (2000)], we consider $L : J^1(\mathbb{R}, M) \rightarrow \mathbb{R}$ to be a smooth Lagrangian function on the 1-jet bundle $J^1(\mathbb{R}, M) \rightarrow \mathbb{R}$, locally expressed by $(t, x^i, v^i) \mapsto L(t, x^i, v^i)$. The so-called *vertical fundamental metrical d -tensor* of L is defined by

$$G_{(i)(j)}^{(1)(1)} = \frac{1}{2} \frac{\partial^2 L}{\partial v^i \partial v^j}. \quad (5.197)$$

Let $h = (h_{11})$ be a semi-Riemannian metric on the temporal manifold \mathbb{R} .

A Lagrangian function $L : J^1(\mathbb{R}, M) \rightarrow \mathbb{R}$ whose vertical fundamental metrical d -tensor is of the form

$$G_{(i)(j)}^{(1)(1)}(t, x^k, v^k) = h^{11}(t) g_{ij}(t, x^k, v^k), \quad (5.198)$$

where $g_{ij}(t, x^k, v^k)$ is a d -tensor on $J^1(\mathbb{R}, M)$, symmetric, of rank n and having a constant signature on $J^1(\mathbb{R}, M)$, is called a Kronecker h -regular Lagrangian function, with respect to the temporal semi-Riemannian metric $h = (h_{11})$.

A pair $\mathbb{R}L^n = (J^1(\mathbb{R}, M), L)$, where $n = \dim M$, which consists of the 1-jet space $J^1(\mathbb{R}, M)$ and a Kronecker h -regular Lagrangian function $L : J^1(\mathbb{R}, M) \rightarrow \mathbb{R}$ is called a *relativistic rheonomic Lagrangian space*.

In our geometrization of the time-dependent Lagrangian function L that we will construct, all entities with geometrical or physical meaning will be directly arisen from the vertical fundamental metrical d -tensor $G_{(i)(j)}^{(1)(1)}$. This fact points out the metrical character (see [Gotay *et. al.* (1998)]) and the naturalness of the subsequent relativistic rheonomic Lagrangian geometry.

For example, suppose that the spatial manifold M is also equipped with a semi-Riemannian metric $g = (g_{ij}(x))$. Then, the time-dependent Lagrangian function $L_1 : J^1(\mathbb{R}, M) \rightarrow \mathbb{R}$ defined by

$$L_1 = h^{11}(t) g_{ij}(x) v^i v^j \quad (5.199)$$

is a Kronecker h -regular time-dependent Lagrangian function. Consequently, the pair $\mathbb{R}L^n = (J^1(\mathbb{R}, M), L_1)$ is a relativistic rheonomic Lagrangian space. We underline that the Lagrangian $L_1 = L_1 \sqrt{|h_{11}|}$ is exactly

the energy Lagrangian whose extremals are the harmonic maps between the semi-Riemannian manifolds (\mathbb{R}, h) and (M, g) . At the same time, this Lagrangian is a basic object in the physical theory of bosonic strings (compare with subsection 6.7 below).

In above notations, taking $U_{(i)}^{(1)}(t, x)$ as a d -tensor-field on $J^1(\mathbb{R}, M)$ and $F : \mathbb{R} \times M \rightarrow \mathbb{R}$ a smooth map, the more general Lagrangian function $L_2 : J^1(\mathbb{R}, M) \rightarrow \mathbb{R}$ defined by

$$L_2 = h^{11}(t)g_{ij}(x)v^i v^j + U_{(i)}^{(1)}(t, x)v^i + F(t, x) \quad (5.200)$$

is also a Kronecker h -regular Lagrangian. The relativistic rheonomic Lagrangian space $\mathbb{R}L^n = (J^1(\mathbb{R}, M), L_2)$ is called the autonomous relativistic rheonomic Lagrangian space of electrodynamics because, in the particular case $h_{11} = 1$, we recover the classical Lagrangian space of electrodynamics [Miron *et. al.* (1988); Miron and Anastasiei (1994)] which governs the movement law of a particle placed concomitantly into a gravitational field and an electromagnetic one. From a physical point of view, the semi-Riemannian metric $h_{11}(t)$ (resp. $g_{ij}(x)$) represents the gravitational potentials of the space \mathbb{R} (resp. M), the d -tensor $U_{(i)}^{(1)}(t, x)$ stands for the electromagnetic potentials and F is a function which is called potential function. The non-dynamical character of spatial gravitational potentials $g_{ij}(x)$ motivates us to use the term of 'autonomous'.

More general, if we consider $g_{ij}(t, x)$ a d -tensor-field on $J^1(\mathbb{R}, M)$, symmetric, of rank n and having a constant signature on $J^1(\mathbb{R}, M)$, we can define the Kronecker h -regular Lagrangian function $L_3 : J^1(\mathbb{R}, M) \rightarrow \mathbb{R}$, setting

$$L_3 = h^{11}(t)g_{ij}(t, x)v^i v^j + U_{(i)}^{(1)}(t, x)v^i + F(t, x). \quad (5.201)$$

The pair $\mathbb{R}L^n = (J^1(\mathbb{R}, M), L_3)$ is a relativistic rheonomic Lagrangian space which is called the non-autonomous relativistic rheonomic Lagrangian space of electrodynamics. Physically, we remark that the gravitational potentials $g_{ij}(t, x)$ of the spatial manifold M are dependent of the temporal coordinate t , emphasizing their dynamical character.

5.7.2 Canonical Nonlinear Connections

Let us consider $h = (h_{11})$ a fixed semi-Riemannian metric on \mathbb{R} and a rheonomic Lagrangian space $\mathbb{R}L^n = (J^1(\mathbb{R}, M), L)$, where L is a Kronecker h -regular Lagrangian function. Let $[a, b] \subset \mathbb{R}$ be a compact interval in

the temporal manifold \mathbb{R} . In this context, we can define the energy action functional of \mathbb{R}^n , setting

$$E : C^\infty(\mathbb{R}, M) \rightarrow \mathbb{R}, \quad E(c) = \int_a^b L(t, x^i, v^i) \sqrt{|h|} dt,$$

where the smooth curve c is locally expressed by $(t) \rightarrow (x^i(t))$ and $v^i = \frac{dx^i}{dt}$.

The extremals of the energy functional E verifies the Euler–Lagrangian equations

$$2G_{(i)(j)}^{(1)(1)} \ddot{x}^j + \frac{\partial^2 L}{\partial x^j \partial v^i} \dot{x}^j - \frac{\partial L}{\partial x^i} + \frac{\partial^2 L}{\partial t \partial v^i} + \frac{\partial L}{\partial v^i} H_{11}^1 = 0, \quad (i = 1, \dots, n), \quad (5.202)$$

where H_{11}^1 are the *Christoffel symbols* of the semi–Riemannian metric h_{11} .

Taking into account the Kronecker h –regularity of the Lagrangian function L , it is possible to rearrange the Euler–Lagrangian equations (5.202) of the Lagrangian $L = L\sqrt{|h|}$, in the Poisson form [Neagu and Udriște (2000a)]

$$\begin{aligned} \Delta_h x^i + 2G^i(t, x^i, v^i) &= 0, \quad (i = 1, \dots, n), \quad \text{where} \quad (5.203) \\ \Delta_h x^i &= h^{11} \{ \ddot{x}^i - H_{11}^1 v^i \}, \quad v^i = \dot{x}^i, \\ 2G^i &= \frac{g^{ii}}{2} \left\{ \frac{\partial^2 L}{\partial x^j \partial v^i} v^j - \frac{\partial L}{\partial x^i} + \frac{\partial^2 L}{\partial t \partial v^i} + \frac{\partial L}{\partial v^i} H_{11}^1 + 2g_{ij} h^{11} H_{11}^1 v^j \right\}. \end{aligned}$$

Denoting $G_{(1)1}^{(r)} = h_{11} G^r$, the geometrical object $G = (G_{(1)1}^{(r)})$ is a spatial spray on $J^1(\mathbb{R}, M)$. By a direct calculation, we deduce that the local geometrical entities of $J^1(\mathbb{R}, M)$

$$\begin{aligned} 2S^k &= \frac{g^{ki}}{2} \left\{ \frac{\partial^2 L}{\partial x^j \partial v^i} v^j - \frac{\partial L}{\partial x^i} \right\}, \\ 2H^k &= \frac{g^{ki}}{2} \left\{ \frac{\partial^2 L}{\partial t \partial v^i} + \frac{\partial L}{\partial v^i} H_{11}^1 \right\}, \quad 2J^k = h^{11} H_{11}^1 v^j, \end{aligned}$$

verify the following transformation rules

$$\begin{aligned} 2S^p &= 2\bar{S}^r \frac{\partial x^p}{\partial \bar{x}^r} + h^{11} \frac{\partial x^p}{\partial \bar{x}^l} \frac{d\bar{t}}{dt} \frac{\partial \bar{x}_\gamma^l}{\partial x^j} v^j, \quad 2H^p = 2\bar{H}^r \frac{\partial x^p}{\partial \bar{x}^r} + h^{11} \frac{\partial x^p}{\partial \bar{x}^l} \frac{d\bar{t}}{dt} \frac{\partial \bar{v}^l}{\partial t}, \\ 2J^p &= 2\bar{J}^r \frac{\partial x^p}{\partial \bar{x}^r} - h^{11} \frac{\partial x^p}{\partial \bar{x}^l} \frac{d\bar{t}}{dt} \frac{\partial \bar{v}^l}{\partial t}. \end{aligned}$$

Consequently, the local entities $2G^p = 2S^p + 2H^p + 2J^p$ can be modified

by the transformation laws

$$2\bar{G}^r = 2G^p \frac{\partial \bar{x}^r}{\partial x^p} - h^{11} \frac{\partial x^p}{\partial \bar{x}^j} \frac{\partial \bar{x}^\mu}{\partial x^p} \bar{v}^j. \quad (5.204)$$

The extremals of the energy functional attached to a Kronecker h -regular Lagrangian function L on $J^1(\mathbb{R}, M)$ are harmonic curves of the time-dependent spray (H, G) , with respect to the semi-Riemannian metric h , defined by the temporal components

$$H_{(1)1}^{(i)} = -\frac{1}{2} H_{11}^1(t) v^i$$

and the local spatial components

$$G_{(1)1}^{(i)} = \frac{h_{11} g^{ik}}{4} \left[\frac{\partial^2 L}{\partial x^j \partial v^k} v^j - \frac{\partial L}{\partial x^k} + \frac{\partial^2 L}{\partial t \partial v^k} + \frac{\partial L}{\partial x^k} H_{11}^1 + 2h^{11} H_{11}^1 g_{kl} v^l \right].$$

The time-dependent spray (H, G) constructed from the previous Theorem is called the canonical time-dependent spray attached to the relativistic rheonomic Lagrangian space $\mathbb{R}L^n$.

In the particular case of an autonomous electrodynamics relativistic rheonomic Lagrangian space (i.e., $g_{ij}(t, x^k, v^k) = g_{ij}(x^k)$), the canonical spatial spray G is given by the components

$$G_{(1)1}^{(i)} = \frac{1}{2} \gamma_{jk}^i v^j v^k + \frac{h_{11} g^{li}}{4} \left[U_{(l)j}^{(1)} v^j + \frac{\partial U_{(l)}^{(1)}}{\partial t} + U_{(l)}^{(1)} H_{11}^1 - \frac{\partial F}{\partial x^l} \right], \quad (5.205)$$

where $U_{(i)j}^{(1)} = \frac{\partial U_{(i)}^{(1)}}{\partial x^j} - \frac{\partial U_{(j)}^{(1)}}{\partial x^i}$.

We have the following Theorem: The pair of local functions $\Gamma = (M_{(1)1}^{(i)}, N_{(1)j}^{(i)})$, which consists of the temporal components

$$M_{(1)1}^{(i)} = 2H_{(1)1}^{(i)} = -H_{11}^1 v^i, \quad (5.206)$$

and the spatial components

$$N_{(1)j}^{(i)} = \frac{\partial G_{(1)1}^{(i)}}{\partial v^j}, \quad (5.207)$$

where $H_{(1)1}^{(i)}$ and $G_{(1)1}^{(i)}$ are the components of the canonical time-dependent spray of $\mathbb{R}L^n$, represents a nonlinear connection on $J^1(\mathbb{R}, M)$.

The nonlinear connection $\Gamma = (M_{(1)1}^{(i)}, N_{(1)j}^{(i)})$ from the preceding Theorem is called the canonical nonlinear connection of the relativistic rheonomic Lagrangian space $\mathbb{R}L^n$.

In the case of an autonomous electrodynamics relativistic rheonomic Lagrangian space (i.e., $g_{ij}(t, x^k, v^k) = g_{ij}(x^k)$), the canonical nonlinear connection becomes $\Gamma = (M_{(1)1}^{(i)}, N_{(1)j}^{(i)})$, where

$$M_{(1)1}^{(i)} = -H_{11}^1 v^i, \quad N_{(1)j}^{(i)} = \gamma_{jk}^i v^k + \frac{h_{11} g^{ik}}{4} U_{(k)j}^{(1)}. \quad (5.208)$$

5.7.3 Cartan's Canonical Connections

The main Theorem of this paper is the Theorem of existence of the Cartan canonical h -normal linear connection CT which allow the subsequent development of the relativistic rheonomic Lagrangian geometry of physical fields, which will be exposed in the next sections.

On the relativistic rheonomic Lagrangian space $\mathbb{R}L^n = (J^1(\mathbb{R}, M), L)$ equipped with its canonical nonlinear connection Γ there is a unique h -normal Γ -linear connection

$$CT = (H_{11}^1, G_{j1}^k, L_{jk}^i, C_{j(k)}^{i(1)})$$

having the metrical properties:

$$(i) \quad g_{ij|k} = 0, \quad g_{ij}|_{(k)}^{(1)} = 0,$$

$$(ii) \quad G_{j1}^k = \frac{g^{ki}}{2} \frac{\delta g_{ij}}{\delta t}, \quad L_{ij}^k = L_{ji}^k, \quad C_{j(k)}^{i(1)} = C_{k(j)}^{i(1)}.$$

To prove this Theorem, let $CT = (\bar{G}_{11}^1, \bar{G}_{j1}^k, \bar{L}_{jk}^i, \bar{C}_{j(k)}^{i(1)})$ be a h -normal Γ -linear connection whose coefficients are defined by $\bar{G}_{11}^1 = H_{11}^1$, $\bar{G}_{j1}^k = \frac{g^{ki}}{2} \frac{\delta g_{ij}}{\delta t}$, and

$$L_{jk}^i = \frac{g^{im}}{2} \left(\frac{\delta g_{jm}}{\delta x^k} + \frac{\delta g_{km}}{\delta x^j} - \frac{\delta g_{jk}}{\delta x^m} \right), \quad C_{j(k)}^{i(1)} = \frac{g^{im}}{2} \left(\frac{\partial g_{jm}}{\partial v^k} + \frac{\partial g_{km}}{\partial v^j} - \frac{\partial g_{jk}}{\partial v^m} \right).$$

By computations, one can verify that CT satisfies the conditions (i) and (ii).

Conversely, let us consider $\bar{CT} = (\bar{G}_{11}^1, \bar{G}_{j1}^k, \bar{L}_{jk}^i, \bar{C}_{j(k)}^{i(1)})$ a h -normal Γ -linear connection which satisfies (i) and (ii). It follows directly that

$$\bar{G}_{11}^1 = H_{11}^1, \quad \text{and} \quad \bar{G}_{j1}^k = \frac{g^{ki}}{2} \frac{\delta g_{ij}}{\delta t}.$$

The condition $g_{ij|k} = 0$ is equivalent with

$$\frac{\delta g_{ij}}{\delta x^k} = g_{mj} \bar{L}_{ik}^m + g_{im} \bar{L}_{jk}^m.$$

Applying the Christoffel process to the indices $\{i, j, k\}$, we find

$$\bar{L}_{jk}^i = \frac{g^{im}}{2} \left(\frac{\delta g_{jm}}{\delta x^k} + \frac{\delta g_{km}}{\delta x^j} - \frac{\delta g_{jk}}{\delta x^m} \right).$$

By analogy, using the relations $C_{j(k)}^{i(1)} = C_{k(j)}^{i(1)}$ and $g_{ij}|_{(k)}^{(1)} = 0$, following a Christoffel process applied to the indices $\{i, j, k\}$, we get

$$\bar{C}_{j(k)}^{i(1)} = \frac{g^{im}}{2} \left(\frac{\partial g_{jm}}{\partial v^k} + \frac{\partial g_{km}}{\partial v^j} - \frac{\partial g_{jk}}{\partial v^m} \right).$$

As a rule, the Cartan canonical connection of a relativistic rheonomic Lagrangian space $\mathbb{R}L^n$ verifies also the properties

$$h_{11/1} = h_{11|k} = h_{11}|_{(k)}^{(1)} = 0 \text{ and } g_{ij/1} = 0. \quad (5.209)$$

The torsion d -tensor T of the Cartan canonical connection of a relativistic rheonomic Lagrangian space is determined by only six local components, because the properties of the Cartan canonical connection imply the relations $T_{ij}^m = 0$ and $S_{(1)(j)(k)}^{(i)(1)(1)} = 0$. At the same time, we point out that the number of the curvature local d -tensors of the Cartan canonical connection not reduces. In conclusion, the curvature d -tensor R of the Cartan canonical connection is determined by five effective local d -tensors. The torsion and curvature d -tensors of the Cartan canonical connection of an $\mathbb{R}L^n$ are called the torsion and curvature of $\mathbb{R}L^n$.

All torsion d -tensors of an autonomous relativistic rheonomic Lagrangian space of electrodynamics vanish, except

$$R_{(1)1j}^{(m)} = -\frac{h_{11}g^{mk}}{4} \left[H_{11}^1 U_{(k)j}^{(1)} + \frac{\partial U_{(k)j}^{(1)}}{\partial t} \right],$$

$$R_{(1)ij}^{(m)} = r_{ijk}^m v^k + \frac{h_{11}g^{mk}}{4} \left[U_{(k)i|j}^{(1)} + U_{(k)j|i}^{(1)} \right],$$

where r_{ijk}^m are the curvature tensors of the semi-Riemannian metric g_{ij} .

5.7.4 General Nonlinear Connections

Recall that a *nonlinear connection* (i.e., a supplementary–horizontal distribution of the vertical distribution of $J^1(\mathbb{R}, M)$) offers the possibility of construction of the vector or covector adapted bases on $J^1(\mathbb{R}, M)$) [Neagu and Udriște (2000a)]. A nonlinear connection Γ on $J^1(\mathbb{R}, M)$ is determined by a pair of local function sets $M_{(1)1}^{(i)}$ and $N_{(1)j}^{(i)}$ which modify by the transformation laws

$$\bar{M}_{(1)1}^{(j)} \frac{d\bar{t}}{dt} = M_{(1)1}^{(k)} \frac{dt}{d\bar{t}} \frac{\partial \bar{x}^j}{\partial x^k} - \frac{\partial \bar{v}^j}{\partial t}, \quad \bar{N}_{(1)k}^{(j)} \frac{\partial \bar{x}^k}{\partial x^i} = N_{(1)i}^{(k)} \frac{dt}{d\bar{t}} \frac{\partial \bar{x}^j}{\partial x^k} - \frac{\partial \bar{v}^j}{\partial x^i}. \quad (5.210)$$

A set of local functions $M_{(1)1}^{(i)}$ (resp. $N_{(1)j}^{(i)}$) on $J^1(\mathbb{R}, M)$, which transform by the rules (5.210) is called a *temporal nonlinear connection* (resp. *spatial nonlinear connection*) on $J^1(\mathbb{R}, M)$.

For example, studying the transformation rules of the local components

$$M_{(1)1}^{(i)} = -H_{11}^1 v^i, \quad N_{(1)j}^{(i)} = \gamma_{jk}^i v^k,$$

where H_{11}^1 (resp. γ_{jk}^i) are the Christoffel symbols of a temporal (resp. spatial) semi-Riemannian metric h (resp. φ), we conclude that $\Gamma_0 = (M_{(1)1}^{(i)}, N_{(1)j}^{(i)})$ represents a nonlinear connection on $J^1(\mathbb{R}, M)$, which is called the canonical nonlinear connection attached to the metric pair (h, φ) .

If $M_{(1)1}^{(i)}$ are the components of a temporal nonlinear connection, then the components $H_{(1)1}^{(i)} = \frac{1}{2} M_{(1)1}^{(i)}$ represent a temporal spray. Conversely, if $H_{(1)1}^{(i)}$ are the components of a temporal spray, then $M_{(1)1}^{(i)} = 2H_{(1)1}^{(i)}$ are the components of a temporal nonlinear connection. If $G_{(1)1}^{(i)}$ are the components of a spatial spray, then the components $N_{(1)j}^{(i)} = \frac{\partial G_{(1)1}^{(i)}}{\partial v^j}$ represent a spatial nonlinear connection.

Conversely, the spatial nonlinear connection $N_{(1)j}^{(i)}$ induces the spatial spray $2G_{(1)1}^{(i)} = N_{(1)j}^{(i)} v^j$.

The previous theorems allow us to conclude that a time-dependent spray (H, G) induces naturally a nonlinear connection Γ on $J^1(\mathbb{R}, M)$, which is called the canonical nonlinear connection associated to the time-dependent spray (H, G) . We point out that the canonical nonlinear connection Γ attached to the time-dependent spray (H, G) is a natural generalization of the canonical nonlinear connection N induced by a time-dependent spray G from the classical rheonomic Lagrangian geometry [Miron *et al.* (1988); Miron and Anastasiei (1994)].

Let $\Gamma = (M_{(1)1}^{(i)} N_{(1)j}^{(i)})$ be a nonlinear connection on $J^1(\mathbb{R}, M)$. Let us consider the geometrical objects,

$$\frac{\delta}{\delta t} = \frac{\partial}{\partial t} - M_{(1)1}^{(j)} \frac{\partial}{\partial v^j}, \quad \frac{\delta}{\delta x^i} = \frac{\partial}{\partial x^i} - N_{(1)i}^{(j)} \frac{\partial}{\partial v^j}, \quad \delta v^i = dy^i + M_{(1)1}^{(i)} dt + N_{(1)j}^{(i)} dx^j.$$

One can deduce that the set of vector-fields $\{\frac{\delta}{\delta t}, \frac{\delta}{\delta x^i}, \frac{\partial}{\partial v^i}\} \subset \mathcal{X}(J^1(\mathbb{R}, M))$ and of covector-fields $\{dt, dx^i, \delta v^i\} \subset \mathcal{X}^*(J^1(\mathbb{R}, M))$ are dual bases. These are called the adapted bases on $J^1(\mathbb{R}, M)$, determined by the nonlinear connection Γ . The big advantage of the adapted bases is that the transformation laws of its elements are simple and natural. The transformation laws of the elements of the adapted bases attached to the nonlinear connection Γ are

$$\begin{aligned} \frac{\delta}{\delta t} &= \frac{d\bar{t}}{dt} \frac{\delta}{\delta \bar{t}}, & \frac{\delta}{\delta x^i} &= \frac{\partial \bar{x}^j}{\partial x^i} \frac{\delta}{\delta \bar{x}^j}, & \frac{\partial}{\partial v^i} &= \frac{\partial \bar{x}^j}{\partial x^i} \frac{dt}{d\bar{t}} \frac{\delta}{\delta \bar{v}^j}, \\ dt &= \frac{d\bar{t}}{dt} d\bar{t}, & dx^i &= \frac{\partial x^i}{\partial \bar{x}^j} d\bar{x}^j, & \delta v^i &= \frac{\partial x^i}{\partial \bar{x}^j} \frac{d\bar{t}}{dt} \delta \bar{v}^j. \end{aligned}$$

5.8 Jets and Action Principles

Recall that in the classical calculus of variations one studies functionals of the form

$$F_L(z) = \int_{\Omega} L(x, z, \nabla z) dx, \quad (\text{with } \Omega \subset \mathbb{R}^n), \quad (5.211)$$

where $x = (x^1, \dots, x^n)$, $dx = dx^1 \wedge \dots \wedge dx^n$, $z = z(x) \in C^1(\bar{\Omega})$, and the *Lagrangian* $L = L(x, z, p)$ is a smooth function of x , z , and $p = (p_1, \dots, p_n)$. The corresponding *Euler-Lagrangian equation*, describing functions $z(x)$ that are *stationary* for such a *functional*, is represented by the second-order PDE [Bryant *et al.* (2003)]

$$\Delta z(x) = F'(z(x)).$$

For example, we may identify a function $z(x)$ with its graph $N \subset \mathbb{R}^{n+1}$, and take the Lagrangian

$$L = \sqrt{1 + \|p\|^2},$$

whose associated functional $F_L(z)$ equals the area of the graph, regarded as a hypersurface in Euclidean space. The Euler-Lagrangian equation describing functions $z(x)$ stationary for this functional is $H = 0$, where H is the *mean curvature* of the graph N .

To study these Lagrangians and Euler–Lagrangian equations geometrically, we have to choose a class of admissible coordinate changes, and there are four natural candidates. In increasing order of generality, they are [Bryant *et al.* (2003)]:

- Classical transformations, of the form $x' = x'(x)$, $z' = z'(z)$; in this situation, we think of (x, z, p) as coordinates on the space $J^1(\mathbb{R}^n, \mathbb{R})$ of 1-jets of maps $\mathbb{R}^n \rightarrow \mathbb{R}$.
- Gauge transformations, of the form $x' = x'(x)$, $z' = z'(x, z)$; here, we think of (x, z, p) as coordinates on the space of 1-jets of sections of a bundle $\mathbb{R}^{n+1} \rightarrow \mathbb{R}^n$, where $x = (x^1, \dots, x^n)$ are coordinates on the base \mathbb{R}^n and $z \in \mathbb{R}$ is a fibre coordinate.
- Point transformations, of the form $x' = x'(x, z)$, $z' = z'(x, z)$; here, we think of (x, z, p) as coordinates on the space of tangent hyperplanes

$$\{dz - p_i dx^i\}^\perp \subset T_{(x^i, z)}(\mathbb{R}^{n+1})$$

of the manifold \mathbb{R}^{n+1} with coordinates (x^1, \dots, x^n, z) .

- Contact transformations, of the form $x' = x'(x, z, p)$, $z' = z'(x, z, p)$, $p' = p'(x, z, p)$, satisfying the equation of differential 1-forms

$$dz' - p'_i dx^{i'} = f \cdot (dz - p_i dx^i)$$

for some function $f(x, z, p) \neq 0$.

Classical calculus of variations primarily concerns the following features of the functional F_L (5.211).

The *first variation* $\delta F_L(z)$ is analogous to the derivative of a function, where $z = z(x)$ is thought of as an independent variable in an infinite-dimensional space of functions. The analog of the condition that a point be critical is the condition that $z(x)$ be stationary for all fixed-boundary variations. Formally, we write

$$\delta F_L(z) = 0,$$

which will give us a second-order scalar PDE for the unknown function $z(x)$ of the form

$$\partial_z L - \partial_{x^i}(\partial_{p_i} L) = 0, \quad (5.212)$$

namely the *Euler–Lagrangian equation* of the Lagrangian $L(x, z, p)$.

In this section we will study the PDE (5.212) in an invariant, geometrical setting, following [Bryant *et al.* (2003)]. As a motivation for this

geometrical approach, we note the fact that Lagrangian is invariant under the large class of *contact transformations*. Also, note that the Lagrangian L determines the functional F_L , but not vice versa. To see this, observe that if we add to $L(x, z, p)$ a *divergence term* and consider

$$L'(x, z, p) = L(x, z, p) + \sum (\partial_{x^i} K^i(x, z) + \partial_z K^i(x, z) p^i)$$

for functions $K^i(x, z)$, then by the *Green's Theorem*, the functionals F_L and $F_{L'}$ differ by a constant depending only on values of z on $\partial\Omega$. L and L' have the same Euler–Lagrangian equations.

Also, there is a relationship between symmetries of a Lagrangian L and conservation laws for the corresponding *Euler–Lagrangian equations*, described by the *Noether Theorem*. A subtlety here is that the group of symmetries of an equivalence class of Lagrangians may be strictly larger than the group of symmetries of any particular representative. We will investigate how this discrepancy is reflected in the space of conservation laws, in a manner that involves global topological issues.

Finally, one considers the second variation $\delta^2 F_L$, analogous to the Hessian of a smooth function, usually with the goal of identifying local minima of the functional. There has been a great deal of analytic work done in this area for classical variational problems, reducing the problem of local minimization to understanding the behavior of certain Jacobi operators, but the geometrical theory is not as well-developed as that of the first variation and the Euler–Lagrangian equations.

Now we turn to *multi-index notation* [Griffiths (1983); Bryant *et al.* (2003); Choquet-Bruhat and DeWitt-Morete (1982)]. An exterior differential system (EDS) is a pair (M, E) consisting of a smooth manifold M and a homogeneous, differentially closed ideal $E \subseteq \Omega^*(M)$ in the algebra of smooth differential forms on M . Some of the EDSs that we study are differentially generated by the sections of a smooth subbundle $I \subseteq T^*M$ of the cotangent bundle of M ; this subbundle, and sometimes its space of sections, is called a *Pfaffian system* on M . It will be useful to use the notation $\{\alpha, \beta, \dots\}$ for the (two-sided) *algebraic ideal* generated by forms α, β, \dots , and to use the notation $\{I\}$ for the algebraic ideal generated by the sections of a Pfaffian system $I \subseteq T^*M$. An *integral manifold* of an EDS (M, E) is a *submanifold immersion* $\iota : N \hookrightarrow M$ for which $\varphi_N \stackrel{\text{def}}{=} \iota^* \varphi = 0$ for all $\varphi \in E$. Integral manifolds of Pfaffian systems are defined similarly.

A differential form φ on the total space of a fibre bundle $\pi : E \rightarrow B$ is said to be *semibasic* if its contraction with any vector-field tangent to

the fibers of π vanishes, or equivalently, if its value at each point $e \in E$ is the pull-back via π_e^* of some form at $\pi(e) \in B$. Some authors call such a form *horizontal*. A stronger condition is that φ be *basic*, meaning that it is locally (in open subsets of E) the pull-back via π^* of a form on the base B [Bryant *et al.* (2003)].

If $(\omega^1, \dots, \omega^n)$ is an ordered basis for a vector space V , then corresponding to a *multi-index* $I = (i_1, \dots, i_k)$ is the k -vector

$$\omega^I = \omega^{i_1} \wedge \dots \wedge \omega^{i_k} \in \wedge^k(V).$$

and for the complete multi-index we define

$$\omega = \omega^1 \wedge \dots \wedge \omega^n.$$

Letting (e_1, \dots, e_n) be a dual basis for V^* , we also define the $(n-k)$ -vector

$$\omega_{(I)} = e_I \lrcorner \omega = e_{i_k} \lrcorner (e_{i_{k-1}} \lrcorner \dots (e_{i_1} \lrcorner \omega) \dots).$$

This $\omega_{(I)}$ is, up to sign, just ω^{I^c} , where I^c is a multi-index complementary to I .

Recall that a *contact manifold* (M, I) is a smooth manifold M of dimension $2n+1$, with a distinguished line subbundle $I \subset T^*M$ of the cotangent bundle which is *non-degenerate* in the sense that for any local 1-form θ generating I ,

$$\theta \wedge (d\theta)^n \neq 0.$$

For example, A 1-jet is an equivalence class of functions having the same value and the same first derivatives at some designated point of the domain. On the space $J^1(\mathbb{R}^n, \mathbb{R})$ of 1-jets of functions, we can take coordinates (x^i, z, p_i) corresponding to the jet at $(x^i) \in \mathbb{R}^n$ of the linear function $f(\bar{x}) = z + p_i(\bar{x}^i - x^i)$. Then we define the *contact form*

$$\theta = dz - p_i dx^i,$$

for which

$$d\theta = -dp_i \wedge dx^i,$$

so the non-degeneracy condition $\theta \wedge (d\theta)^n \neq 0$ is apparent. In fact, the *Pfaff Theorem* [Bryant *et al.* (2003)] implies that every contact manifold is locally isomorphic to this example; that is, every contact manifold (M, I) has local coordinates (x^i, z, p_i) for which the form $\theta = dz - p_i dx^i$ generates I .

Let (M, I) be a contact manifold of dimension $2n + 1$, and assume that I is generated by a global, non-vanishing section $\theta \in \Gamma(I)$; this assumption only simplifies our notation, and would in any case hold on a double-cover of M . Sections of I generate the *contact differential ideal*

$$I = \{\theta, d\theta\} \subset \Omega^*(M)$$

in the exterior algebra of differential forms on M . A *Legendre submanifold* of M is an immersion $\iota : N \hookrightarrow M$ of an n D submanifold N such that $\iota^*\theta = 0$ for any contact form $\theta \in \Gamma(I)$; in this case $\iota^*d\theta = 0$ as well, so a Legendre submanifold is the same thing as an integral manifold of the differential ideal I . In Pfaff coordinates with $\theta = dz - p_i dx^i$, one such integral manifold is $N_0 = \{z = p_i = 0\}$. To see other Legendre submanifolds ‘near’ this one, note that any submanifold C^1 -close to N_0 satisfies the *independence condition* [Bryant *et al.* (2003)]

$$dx^1 \wedge \cdots \wedge dx^n \neq 0,$$

and can therefore be described locally as a graph

$$N = \{(x^i, z(x), p_i(x))\}.$$

In this case, we have

$$\theta|_N = 0 \quad \text{iff} \quad p_i(x) = \partial_{x^i} z(x).$$

Therefore, N is determined by the function $z(x)$, and conversely, every function $z(x)$ determines such an N ; we informally say that ‘the generic Legendre submanifold depends locally on one arbitrary function of n variables’. Legendre submanifolds of this form, with $dx|_N \neq 0$, are called *transverse*.

Now, we are interested in functionals given by triples (M, I, Λ) , where (M, I) is a $(2n + 1)$ D contact manifold, and $\Lambda \in \Omega^n(M)$ is a differential form of degree n on M ; such a Λ will be referred to as a *Lagrangian* on (M, I) [Bryant *et al.* (2003)]. We then define a functional on the set of smooth, compact Legendre submanifolds $N \subset M$, possibly with boundary ∂N , by

$$F_\Lambda(N) = \int_N \Lambda.$$

The classical variational problems described above may be recovered from this notion by taking $M = J^1(\mathbb{R}^n, \mathbb{R}) \cong \mathbb{R}^{2n+1}$ with coordinates (x^i, z, p_i) , I generated by $\theta = dz - p_i dx^i$, and $\Lambda = L(x^i, z, p_i) dx$. This formulation

also admits certain functionals depending on second derivatives of $z(x)$, because there may be dp_i -terms in Λ . Later, we will restrict attention to a class of functionals which, possibly after a contact transformation, can be expressed without second derivatives.

Suppose given a Lagrangian $\Lambda \in \Omega^n(M)$ on a *contact manifold* (M, I) , and a fixed-boundary variation of Legendre submanifold $F : N \times [0, 1] \rightarrow M$; we wish to calculate $\frac{d}{dt}(\int_{N_t} \Lambda)$.

To do this, first recall the calculation of the *Poincaré–Cartan form* for the equivalence class $[\Lambda] \in \bar{H}^n$. Because $I^{n+1} = \Omega^{n+1}(M)$, we can write [Bryant *et al.* (2003)]

$$d\Lambda = \theta \wedge \alpha + d\theta \wedge \beta = \theta \wedge (\alpha + d\beta) + d(\theta \wedge \beta),$$

and then

$$\Pi = \theta \wedge (\alpha + d\beta) = d(\Lambda - \theta \wedge \beta). \quad (5.213)$$

We are looking for conditions on a *Legendre submanifold* $f : N \hookrightarrow M$ to be *stationary* for $[\Lambda]$ under all fixed-boundary variations, in the sense that $\frac{d}{dt}|_{t=0}(\int_{N_t} \Lambda) = 0$ whenever $F|_{t=0} = f$. We calculate

$$\partial_t \int_{N_t} \Lambda = \partial_t \int_{N_t} (\Lambda - \theta \wedge \beta) = \int_{N_t} L_{\partial_t}(\Lambda - \theta \wedge \beta) = \int_{N_t} \partial_t \lrcorner \Pi.$$

One might express this result as

$$\delta(F_\Lambda)_N(v) = \text{int}_{N_t} v \lrcorner f^* \Pi,$$

where the variational vector-field v , lying in the space $\Gamma_0(f^*TM)$ of sections of f^*TM vanishing along ∂N , plays the role of ∂_t . The condition $\Pi \equiv 0 \pmod{\{I\}}$ allows us to write $\Pi = \theta \wedge \Psi$ for some n -form Ψ , not uniquely determined, and we have [Bryant *et al.* (2003)]

$$\frac{d}{dt} \Big|_{t=0} \int_{N_t} \Lambda = \int_N g f^* \Psi,$$

where $g = (\partial_t \lrcorner F^* \theta)|_{t=0}$. It was shown previously that this g could locally be chosen arbitrarily in the interior N° , so the necessary and sufficient condition for a Legendre submanifold $f : N \hookrightarrow M$ to be stationary for F_Λ is that $f^* \Psi = 0$.

In the particular classical situation where $M = \{(x^i, z, p_i)\}$, $\theta = dz - p_i dx^i$, and $\Lambda = L(x, z, p) dx$, we have

$$d\Lambda = L_z \theta \wedge dx + L_{p_i} dp_i \wedge dx = \theta \wedge L_z dx - d\theta \wedge L_{p_i} dx_{(i)},$$

so referring to (5.213),

$$\Pi = \theta \wedge (L_z dx - d(L_{p_i} dx_{(i)})) = \theta \wedge \Psi.$$

Now, for a transverse Legendre submanifold $N = \{(x^i, z(x), z_{x^i}(x))\}$, we have $\Psi|_N = 0$ iff (5.212) is valid along N .

Later, (see section 5.10 below) we will extend the jet-action formalism presented here – to the rigorous (and elegant) jet formulation of path-integrals in physical field systems.

5.9 Application: Jets and Lagrangian Field Theory

In this subsection we will apply the jet formalism defined in subsection 4.14.12.5 above, and already applied for development of the time-dependent mechanics in subsection 5.6.1 above, to formulate the *first-order Lagrangian field theory* on fibre bundles (see [Sardanashvily (1993); Sardanashvily (1995); Giachetta *et. al.* (1997); Mangiarotti and Sardanashvily (2000a); Sardanashvily (2002a)] for details).

Recall that the configuration space of the first-order Lagrangian field theory on a fibre bundle $Y \rightarrow X$, coordinated by $(x^\alpha, y^i, y_\alpha^i)$, is the 1-jet space $J^1(X, Y)$ of the bundle $Y \rightarrow X$, coordinated by $(x^\alpha, y^i, y_\alpha^i)$. Therefore, a first-order Lagrangian $L : J^1(X, Y) \rightarrow \wedge^n T^*X$ is defined as a *horizontal density* on $J^1(X, Y)$,

$$L = \mathcal{L}(x^\alpha, y^i, y_\alpha^i)\omega, \quad \text{with} \quad \omega = dx^1 \wedge \dots \wedge dx^n, \quad (n = \dim X). \quad (5.214)$$

Let us follow the standard formulation of the variational problem on fibre bundles where deformations of sections of a fibre bundle $Y \rightarrow X$ are induced by local 1-parameter groups of automorphisms of $Y \rightarrow X$ over X (the so-called vertical gauge transformations). Here, we will not study the calculus of variations in depth, but apply in a straightforward manner the *first variational formula* (5.109) (for technical details, see [Sardanashvily (1993); Sardanashvily (1995); Giachetta *et. al.* (1997); Mangiarotti and Sardanashvily (2000a); Sardanashvily (2002a)]).

Recall that a projectable vector-field u on a fibre bundle $Y \rightarrow X$ is an infinitesimal generator of a local 1-parameter group of gauge transformations of $Y \rightarrow X$. Therefore, one can think of its jet prolongation j^1u (5.9) as being the infinitesimal generator of gauge transformations of the configuration space $J^1(X, Y)$. Let the Lie derivative of a Lagrangian L along j^1u

be given by

$$\mathfrak{L}_{j^1 u} L = [\partial_\alpha u^\alpha \mathcal{L} + (u^\alpha \partial_\alpha + u^i \partial_i + (d_\alpha u^i - y_\mu^i \partial_\alpha u^\mu) \partial_i^\alpha) \mathcal{L}] \omega. \quad (5.215)$$

The *first variational formula* (5.109) gives its *canonical decomposition* (in accordance with the general variational problem), which reads

$$\begin{aligned} \mathfrak{L}_{j^1 u} L &= u_V \rfloor \mathcal{E}_L + d_H h_0(u) \rfloor H_L \\ &= (u^i - y_\mu^i u^\mu) (\partial_i - d_\alpha \partial_i^\alpha) \mathcal{L} \omega - d_\alpha [\pi_i^\alpha (u^\mu y_\mu^i - u^i) - u^\alpha \mathcal{L}] \omega. \end{aligned} \quad (5.216)$$

In the canonical decomposition (5.216), $u_V = (u) \rfloor \theta^i \partial_i$; the map

$$\mathcal{E}_L : J^2(X, Y) \rightarrow T^*Y \wedge (\wedge^n T^*X), \quad \text{given by} \quad \mathcal{E}_L = (\partial_i \mathcal{L} - d_\alpha \pi_i^\alpha) \theta^i \wedge \omega, \quad (5.217)$$

(with $\pi_i^\alpha = \partial_i^\alpha \mathcal{L}$) is the *Euler–Lagrangian operator* associated to the Lagrangian L ; and the map

$$H_L : J^1(X, Y) \rightarrow M_Y = T^*Y \wedge (\wedge^{n-1} T^*X), \quad \text{given by} \quad (5.218)$$

$$H_L = L + \pi_i^\alpha \theta^i \wedge \omega_\alpha = \pi_i^\alpha dy^i \wedge \omega_\alpha + (\mathcal{L} - \pi_i^\alpha y_\alpha^i) \omega, \quad (5.219)$$

is called the *Poincaré–Cartan form*.

The *kernel* of the Euler–Lagrangian operator \mathcal{E}_L (5.217) defines the system of *second-order Euler–Lagrangian equations*, in local coordinates given by

$$(\partial_i - d_\alpha \partial_i^\alpha) \mathcal{L} = 0, \quad (5.220)$$

A solution of these equations is a section $s : X \rightarrow Y$ of the fibre bundle $Y \rightarrow X$, whose second-order jet prolongation $j^2 s$ lives in (5.220), i.e.,

$$\partial_i \mathcal{L} \circ s - (\partial_\alpha + \partial_\alpha s^j \partial_j + \partial_\alpha \partial_\mu s^j \partial_j^\mu) \partial_i^\alpha \mathcal{L} \circ s = 0. \quad (5.221)$$

Different Lagrangians L and L' can lead to the same Euler–Lagrangian operator \mathcal{E}_L if their difference $L_0 = L - L'$ is a variationally trivial Lagrangian, whose Euler–Lagrangian operator vanishes identically. A Lagrangian L_0 is called *variationally trivial* iff

$$L_0 = h_0(\varphi), \quad (5.222)$$

where φ is a closed n -form on Y . We have at least locally $\varphi = d\xi$, and then

$$L_0 = h_0(d\xi) = d_H(h_0(\xi)) = d_\alpha h_0(\xi)^\alpha \omega, \quad h_0(\xi) = h_0(\xi)^\alpha \omega_\alpha.$$

The Poincaré–Cartan form H_L (5.218) is called a *Lepagean equivalent* of a Lagrangian L if $h_0(H_L) = L$. In contrast with other Lepagean forms (see [Giachetta *et. al.* (1997); Mangiarotti and Sardanashvily (2000a)]), H_L is a horizontal form on the affine jet bundle $J^1(X, Y) \rightarrow Y$.

The fibre bundle $M_Y = T^*Y \wedge (\wedge^{n-1}T^*X)$, figuring in the Poincaré–Cartan form (5.218) is called the *homogeneous Legendre bundle*. It has holonomic local coordinates $(x^\alpha, y^i, p_i^\alpha, p)$ with transition functions

$$p_i'^\alpha = \det\left(\frac{\partial x^\varepsilon}{\partial x'^\nu}\right) \frac{\partial y^j}{\partial y'^i} \frac{\partial x'^\alpha}{\partial x^\mu} p_j^\mu, \quad p' = \det\left(\frac{\partial x^\varepsilon}{\partial x'^\nu}\right) \left(p - \frac{\partial y^j}{\partial y'^i} \frac{\partial y'^i}{\partial x^\mu} p_j^\mu\right). \quad (5.223)$$

Relative to these coordinates, the map (5.218) reads

$$(p_i^\mu, p) \circ H_L = (\pi_i^\mu, \mathcal{L} - \pi_i^\mu y_\mu^i).$$

The transition functions (5.223) shows that M_Y is a 1D affine bundle

$$\pi_{M\Pi} : M_Y \rightarrow \Pi \quad (5.224)$$

over the *Legendre bundle*

$$\Pi = \wedge^n T^*X \otimes V^*Y \otimes TX = V^*Y \wedge (\wedge^{n-1}T^*X), \quad (5.225)$$

with holonomic coordinates $(x^\alpha, y^i, p_i^\alpha)$. Then the composition

$$\widehat{L} = \pi_{M\Pi} \circ H_L : J^1(X, Y) \rightarrow \Pi, \quad (x^\alpha, y^i, p_i^\alpha) \circ \widehat{L} = (x^\alpha, y^i, \pi_i^\alpha), \quad (5.226)$$

is the well-known *Legendre map*. One can think of p_i^α as being the covariant momenta of field functions, and the Legendre bundle Π (5.225) plays the role of a finite-dimensional *momentum phase-space of fields* in the covariant Hamiltonian field theory (see subsection 5.10 below).

The first variational formula (5.216) gives the standard procedure for the study of differential conservation laws in Lagrangian field theory as follows.

Let u be a projectable vector-field on a fibre bundle $Y \rightarrow X$ treated as the infinitesimal generator of a local 1-parameter group G_u of gauge transformations. On-shell, i.e., on the kernel (5.220) of the Euler–Lagrangian operator \mathcal{E}_L , the first variational formula (5.216) leads to the *weak identity*

$$\mathfrak{L}_{j^1 u} L \approx -d_\alpha \mathfrak{J}^\alpha \omega, \quad \text{where} \quad (5.227)$$

$$\mathfrak{J} = \mathfrak{J}^\alpha \omega_\alpha, \quad \mathfrak{J}^\alpha = \pi_i^\alpha (u^\mu y_\mu^i - u^i) - u^\alpha \mathcal{L}, \quad (5.228)$$

is the *symmetry current* along the vector-field u . Let a Lagrangian L be invariant under the gauge group G_u . This implies that the Lie derivative $\mathcal{L}_{j^1 u} L$ (5.215) vanishes. Then we get the *weak conservation law*

$$d_\alpha \mathfrak{J}^\alpha \approx 0 \quad (5.229)$$

of the symmetry current \mathfrak{J} (5.228).²

The weak conservation law (5.229) leads to the differential conservation law

$$\partial_\alpha (\mathfrak{J}^\alpha \circ s) = 0 \quad (5.230)$$

on solutions $s : X \rightarrow Y$ (5.221) of the Euler–Lagrangian equations (5.220). It implies the *integral conservation law*

$$\int_{\partial N} s^* \mathfrak{J} = 0, \quad (5.231)$$

where N is a compact n D submanifold of X with the boundary ∂N .

In gauge theory, the symmetry current \mathfrak{J} (5.228) takes the form

$$\mathfrak{J} = W + d_H U = (W^\alpha + d_\mu U^{\mu\alpha}) \omega_\alpha, \quad (5.232)$$

where the term W depends only on the *variational derivatives*

$$\delta_i \mathcal{L} = (\partial_i - d_\alpha \partial_i^\alpha) \mathcal{L}, \quad (5.233)$$

i.e., $W \approx 0$. The tensor-field $U = U^{\mu\alpha} \omega_{\mu\alpha} : J^1(X, Y) \rightarrow \wedge^{n-2} T^*X$ is a horizontal $(n-2)$ –form on $J^1(X, Y) \rightarrow X$. Then one says that \mathfrak{J} *reduces to the superpotential* U (see [Fatibene *et. al.* (1994); Giachetta *et. al.* (1997); Sardanashvily (1997)]). On-shell, such a symmetry current reduces to a d_H –exact form (5.232). In this way, the differential conservation law (5.230) and the integral conservation law (5.231) become tautological. At the same time, the superpotential form (5.232) of \mathfrak{J} implies the following integral relation

$$\int_{N^{n-1}} s^* \mathfrak{J} = \int_{\partial N^{n-1}} s^* U, \quad (5.234)$$

²The first variational formula defines the symmetry current (5.228) modulo the terms $d_\mu (c_i^{\mu\alpha} (y_\nu^i u^\nu - u^i))$, where $c_i^{\mu\alpha}$ are arbitrary skew-symmetric functions on Y [Giachetta *et. al.* (1997)]. Here, we set aside these boundary terms which are independent of a Lagrangian L .

where N^{n-1} is a compact oriented $(n-1)$ D submanifold of X with the boundary ∂N^{n-1} . One can think of this relation as being a part of the Euler–Lagrangian equations written in an integral form.

Let us consider conservation laws in the case of gauge transformations which preserve the Euler–Lagrangian operator \mathcal{E}_L , but not necessarily a Lagrangian L . Let u be a projectable vector-field on $Y \rightarrow X$, which is the infinitesimal generator of a local 1-parameter group of such transformations, i.e.,

$$\mathfrak{L}_{j^2 u} \mathcal{E}_L = 0,$$

where $j^2 u$ is the second-order jet prolongation of the vector-field u . There is the useful relation [Giachetta *et. al.* (1997)]

$$\mathfrak{L}_{j^2 u} \mathcal{E}_L = \mathcal{E}_{\mathfrak{L}_{j^1 u} L}. \quad (5.235)$$

Then, in accordance with (5.222), we have locally

$$\mathfrak{L}_{j^1 u} L = d_\alpha h_0(\xi)^\alpha \omega. \quad (5.236)$$

In this case, the weak identity (5.227) reads

$$d_\alpha (h_0(\xi)^\alpha - \mathfrak{J}^\alpha) \approx 0, \quad (5.237)$$

where \mathfrak{J} is the symmetry current (5.228).

Background fields, which do not live in the dynamical shell (5.220), violate conservation laws as follows. Let us consider the product

$$Y_{tot} = Y \times Y' \quad (5.238)$$

of a fibre bundle $Y \rightarrow X$, coordinated by (x^α, y^i) , whose sections are dynamical fields and of a fibre bundle $Y' \rightarrow X$, coordinated by (x^α, y^A) , whose sections are *background fields* that take the background values

$$y^B = \phi^B(x), \quad \text{and} \quad y_\alpha^B = \partial_\alpha \phi^B(x).$$

A Lagrangian L of dynamical and background fields is defined on the total configuration space $J^1(X, Y)_{tot}$. Let u be a projectable vector-field on Y_{tot} which also projects onto Y' because gauge transformations of background fields do not depend on dynamical fields. This vector-field takes the coordinate form

$$u = u^\alpha(x^\mu) \partial_\alpha + u^A(x^\mu, y^B) \partial_A + u^i(x^\mu, y^B, y^j) \partial_i. \quad (5.239)$$

Substitution of u (5.239) in the formula (5.216) leads to the first variational formula in the presence of background fields:

$$\begin{aligned} \partial_\alpha u^\alpha \mathcal{L} + [u^\alpha \partial_\alpha + u^A \partial_A + u^i \partial_i + (d_\alpha u^A - y_\mu^A \partial_\alpha u^\mu) \partial_A^\alpha] \mathcal{L} = & (u^A - y_\alpha^A u^\alpha) \partial_A \mathcal{L} + \pi_A^\alpha d_\alpha (u^A - y_\mu^A u^\mu) \\ & + (u^i - y_\alpha^i u^\alpha) \delta_i \mathcal{L} - d_\alpha [\pi_i^\alpha (u^\mu y_\mu^i - u^i) - u^\alpha \mathcal{L}] \end{aligned} \quad (5.241)$$

Then we have on the shell (5.220) the weak identity

$$\begin{aligned} \partial_\alpha u^\alpha \mathcal{L} + [u^\alpha \partial_\alpha + u^A \partial_A + u^i \partial_i + (d_\alpha u^A - y_\mu^A \partial_\alpha u^\mu) \partial_A^\alpha + (d_\alpha u^i - y_\mu^i \partial_\alpha u^\mu) \partial_i^\alpha] \mathcal{L} \\ \approx (u^A - y_\alpha^A u^\alpha) \partial_A \mathcal{L} + \pi_A^\alpha d_\alpha (u^A - y_\mu^A u^\mu) - d_\alpha [\pi_i^\alpha (u^\mu y_\mu^i - u^i) - u^\alpha \mathcal{L}]. \end{aligned}$$

If a total Lagrangian L is invariant under gauge transformations of Y_{tot} , we get the weak identity

$$(u^A - y_\mu^A u^\mu) \partial_A \mathcal{L} + \pi_A^\alpha d_\alpha (u^A - y_\mu^A u^\mu) \approx d_\alpha \mathfrak{J}^\alpha, \quad (5.242)$$

which is the transformation law of the symmetry current \mathfrak{J} in the presence of background fields.

5.9.1 Lagrangian Conservation Laws

In the first-order Lagrangian field theory, we have the following differential transformation and conservation laws on solutions $s : X \rightarrow Y$ (5.221) of the Euler–Lagrangian equations (5.220).

Recall that given fibre coordinates (x^α, y^i) of Y , the jet space $J^1(X, Y)$ is equipped with the adapted coordinates $(x^\alpha, y^i, y_\alpha^i)$, while the first-order *Lagrangian density* on $J^1(X, Y)$ is defined as the map

$$\begin{aligned} L : J^1(X, Y) &\rightarrow \wedge^n T^*X, \quad (n = \dim X), \\ L &= \mathcal{L}(x^\alpha, y^i, y_\alpha^i) \omega, \quad \text{with} \quad \omega = dx^1 \wedge \dots \wedge dx^n. \end{aligned}$$

The corresponding *first-order Euler–Lagrangian equations* for sections $\bar{s} : X \rightarrow J^1(X, Y)$ of the jet bundle $J^1(X, Y) \rightarrow X$ read

$$\partial_\alpha \bar{s}^i = \bar{s}_\alpha^i, \quad \partial_i \mathcal{L} - (\partial_\alpha + \bar{s}_\alpha^j \partial_j + \partial_\alpha \bar{s}_\alpha^j \partial_j^\alpha) \partial_i^\alpha \mathcal{L} = 0. \quad (5.243)$$

We consider the Lie derivatives of Lagrangian densities in order to get differential conservation laws. Let

$$u = u^\alpha(x) \partial_\alpha + u^i(y) \partial_i$$

be a projectable vector-field on $Y \rightarrow X$ and \bar{u} its jet lift (5.9) onto $J^1(X, Y) \rightarrow X$. Given L , let us compute the Lie derivative $\mathfrak{L}_{\bar{u}}L$. We get the identity

$$\bar{s}^* \mathfrak{L}_{\bar{u}}L \approx -\frac{d}{dx^\alpha} [\pi_i^\alpha (u^\alpha \bar{s}_\alpha^i - u^i) - u^\alpha \mathcal{L}] \omega, \quad \pi_i^\alpha = \partial_i^\alpha \mathcal{L}, \quad (5.244)$$

modulo the Euler–Lagrangian equations (5.243).

Let L be a Lagrangian density on the jet space $J^1(X, Y)$. For the sake of simplicity, we shall denote the pull-back π_0^*L of L onto $J^2(X, Y)$ by the same symbol L .

Let u be a projectable vector-field on $Y \rightarrow X$ and \bar{u} its jet lift (5.9) onto the configuration bundle $J^1(X, Y) \rightarrow X$. Recall that the vector-field u is associated with some 1-parameter group of transformations of Y .

Let us calculate the Lie derivative $\mathfrak{L}_{\bar{u}}L$ of the horizontal density L when its Lepagean equivalent is chosen to be the Poincaré–Cartan form Ξ_L , given by the coordinate expression

$$\Xi_L = L\omega + \pi_i^\alpha (dy^i - y_\alpha^i dx^\alpha) \wedge \omega_\alpha. \quad (5.245)$$

In this case we recover the first variational formula (5.216) for projectable vector-fields on Y as (see [Giachetta and Mangiarotti (1990); Sardanashvily (1997)])

$$\mathfrak{L}_{\bar{u}}L = u_V \rfloor \mathcal{E}_L + h_0(d\bar{u} \rfloor \Xi_L). \quad (5.246)$$

Since the Poincaré–Cartan form Ξ_L is a horizontal form on the jet bundle $J^1(X, Y) \rightarrow Y$, the formula (5.246) takes the form

$$\mathfrak{L}_{\bar{u}}L = u_V \rfloor \mathcal{E}_L + d_H h_0(u \rfloor \Xi_L). \quad (5.247)$$

Being restricted to the kernel

$$[\partial_i - (\partial_\alpha + y_\alpha^j \partial_j + y_{\alpha\lambda}^j \partial_j^\lambda) \partial_i^\alpha] \mathcal{L} = 0$$

of the Euler–Lagrangian operator \mathcal{E}_L (5.217), the equality (5.247) reduces to the weak identity

$$\mathfrak{L}_{\bar{u}}L \approx d_H h_0(u \rfloor \Xi_L), \quad (5.248)$$

$$\partial_\alpha u^\alpha \mathcal{L} + [u^\alpha \partial_\alpha + u^i \partial_i + (\partial_\alpha u^i + y_\alpha^j \partial_j u^i - y_\alpha^i \partial_\alpha u^\alpha) \partial_i^\alpha] \mathcal{L} \approx \widehat{\partial}_\alpha [\pi_i^\alpha (u^i - u^\alpha y_\alpha^i) + u^\alpha \mathcal{L}],$$

$$\widehat{\partial}_\alpha = \partial_\alpha + y_\alpha^i \partial_i + y_{\alpha\lambda}^i \partial_i^\lambda.$$

On solutions s of the Euler–Lagrangian equations, the weak identity (5.248) becomes the weak differential transformation law

$$s^* \mathfrak{L}_{\bar{u}} L \approx d(s^* u] \Xi_L) \quad (5.249)$$

which takes the coordinate form (5.244).

Note that, in order to get the differential transformation laws on solutions s of a given system of Euler–Lagrangian equations, one can examine other Lepagean equivalents ρ_L of the Lagrangian density L , besides the Poincaré–Cartan form Ξ_L . In this case, the first variational formula (5.246) and the corresponding weak identity

$$\mathfrak{L}_{\bar{u}} L \approx h_0(d\bar{u}] \rho_L)$$

differ from relations (5.246) and (5.248) respectively in the strong identity

$$0 = h_0(d\bar{u}] \varepsilon) = d_H h_0(\bar{u}] \varepsilon), \quad (5.250)$$

where $\rho_L = \Xi_L + \varepsilon$. From the physical point of view, it means that different Lepagean equivalents result in different superpotentials $h_0(\bar{u}] \varepsilon)$.

The form ε in the identity (5.250) has the coordinate expression

$$\varepsilon = -(\hat{\partial}_\nu c_i^{\alpha\nu} \hat{d}y^i + c_i^{\alpha\nu} \hat{d}y_\nu^i) \wedge \omega_\alpha + \chi.$$

It is the general local expression for Lepagean equivalents of the zero Lagrangian density. We have

$$h_0(\bar{u}] \varepsilon) = \hat{\partial}_\nu [(u^i - y_\alpha^i u^\alpha) c_i^{\alpha\nu}] \omega_\alpha.$$

One can consider also other Lagrangian densities L' which possess the same Euler–Lagrangian operator \mathcal{E}_L . Then the first variational formula and the corresponding weak identity differ from relations (5.246) and (5.248) respectively in the strong identity

$$\mathfrak{L}_{\bar{u}} h_0(\epsilon) = h_0(d\bar{u}] \epsilon) \quad (5.251)$$

where ϵ is some closed exterior form on Y . However, if the form $h_0(\epsilon)$ possesses the same symmetries as the Lagrangian density L only, the contribution of the strong identity (5.251) into the weak identity (5.248) is not tautological.

Note that the weak identity (5.248) is linear in the vector-field u , and we can consider superposition of different weak identities (5.248) corresponding

to different vector-fields u . For example, if u and u' are projectable vector-fields on the bundle $Y \rightarrow X$ which are projected onto the same vector-field on the base X , their difference $u - u'$ is a vertical vector-field on $Y \rightarrow X$. Therefore, the difference of the weak identity (5.248) with respect to the vector-fields u and u' results in the weak identity (5.248) with respect to the vertical vector-field $u - u'$.

Now let us consider the case when a Lagrangian density L depends on background fields. We define such a Lagrangian density as the pull-back of the Lagrangian density L_{tot} on the total configuration space by some fixed sections $\phi(x)$ describing background fields.

Let us again consider the product (5.238), namely $Y_{tot} = Y \times Y'$, of the bundle Y whose sections are *dynamical fields* and the bundle Y' whose sections ϕ play the role of *background fields*. Let the bundles Y and Y' be coordinated by (x^α, y^i) and (x^α, y^A) respectively. The Lagrangian density L_{tot} is defined on the total configuration space $J^1(X, Y)_{tot}$.

Let u be a projectable vector-field on Y_{tot} which is also projectable with respect to projection $Y \times Y' \rightarrow Y'$. It has the coordinate form

$$u = u^\alpha(x) \partial_\alpha + u^A(x^\alpha, y^B) \partial_A + u^i(x^\alpha, y^B, y^j) \partial_i,$$

showing that transformations of background fields are independent on dynamical fields.

Calculating the Lie derivative of the Lagrangian density L_{tot} by this vector-field, we get the equality

$$\begin{aligned} & \partial_\alpha u^\alpha \mathcal{L}_{tot} + [u^\alpha \partial_\alpha + u^A \partial_A + u^i \partial_i + (\partial_\alpha u^A + y_\alpha^B \partial_B u^A - y_\alpha^A \partial_\alpha u^\alpha) \partial_A^\alpha \\ & + (\partial_\alpha u^i + y_\alpha^B \partial_B u^i + y_\alpha^j \partial_j u^i - y_\alpha^i \partial_\alpha u^\alpha) \partial_i^\alpha] \mathcal{L}_{tot} = \widehat{\partial}_\alpha [\pi_i^\alpha (u^i - u^\alpha y_\alpha^i) + u^\alpha \mathcal{L}_{tot}] \\ & + (u^i - y_\alpha^i u^\alpha) (\partial_i - \widehat{\partial}_i^\alpha) \mathcal{L}_{tot} + (u^A - y_\alpha^A u^\alpha) \partial_A \mathcal{L}_{tot} + \pi_A^\alpha \widehat{\partial}_\alpha (u^A - y_\alpha^A u^\alpha), \end{aligned}$$

which can be rewritten as

$$\begin{aligned} & \partial_\alpha u^\alpha \mathcal{L}_{tot} + [u^\alpha (\partial_\alpha + y_\alpha^B \partial_B + y_{\alpha\lambda}^B \partial_B^\alpha) \mathcal{L} + u^i \partial_i + (\widehat{\partial}_\alpha u^i - y_\alpha^i \partial_\alpha u^\alpha) \partial_i^\alpha] \mathcal{L}_{tot} \\ & = \widehat{\partial}_\alpha [\pi_i^\alpha (u^i - u^\alpha y_\alpha^i) + u^\alpha \mathcal{L}_{tot}] + (u^i - y_\alpha^i u^\alpha) (\partial_i - \widehat{\partial}_i^\alpha) \mathcal{L}_{tot}. \end{aligned}$$

The pull-back of this equality to the bundle $Y \rightarrow X$ by sections $\phi^A(x)$ of the bundle Y' which describe the background fields results in the familiar expression (5.247) and the familiar weak identity (5.248) for the Lagrangian density $L = \phi^* L_{tot}$. Now the partial derivative ∂_α can be written as

$$\partial_\alpha = \widetilde{\partial}_\alpha + \partial_\alpha \phi^B \partial_B + \partial_\alpha \partial_\alpha \phi^B \partial_B^\alpha,$$

where $\tilde{\partial}_\alpha$ denote the partial derivatives with respect to the coordinates x^α on which the Lagrangian density L_{tot} depends explicitly.

Note that Lagrangian densities of field models almost never depend explicitly on the world coordinates x^α . At the same time, almost all field models describe fields in the presence of a background world metric g on the base manifold X , except topological field theories whose classical Lagrangian densities are independent on g [Birmingham *et. al.* (1991)] and the gravitation theory where a world metric g is a dynamical field.

By a world metric on X is denoted a nondegenerate fibre metric $g^{\alpha\nu}$ in cotangent and tangent bundles of X . In this case, the partial derivative $\partial_\alpha \mathcal{L}$ in the weak identity (5.248) contains the term $\frac{\partial \mathcal{L}}{\partial g^{\alpha\nu}} \partial_\alpha g^{\alpha\nu}$, so that the metric *stress-energy-momentum tensor* of fields (SEM-tensor, for short, see subsection 5.12.1 below)

$$t_{\alpha\nu} \sqrt{|g|} = 2 \frac{\partial \mathcal{L}}{\partial g^{\alpha\nu}}, \quad |g| = |\det(g_{\alpha\nu})|.$$

is called into play.

The weak identity (5.248) and the weak transformation law (5.249) are basic for our analysis of differential transformation and conservation laws in field theory.

In particular, one says that an isomorphism Φ of the fibre bundle $Y \rightarrow X$ is an invariant transformation if its jet prolongation $j^1\Phi$ preserves the Lagrangian density L , i.e.,

$$j^{1*}\Phi L = L.$$

Let u be a projectable vector-field on $Y \rightarrow X$. The corresponding local 1-parameter groups of isomorphisms of Y are invariant transformations iff the *strong equality*: $\mathfrak{L}_{\bar{u}}L = 0$ holds. In this case, we have the corresponding weak conservation law

$$d(s^*u] \Xi_L) \approx 0. \quad (5.252)$$

An isomorphism Φ of the bundle $Y \rightarrow X$ is called the generalized invariant transformation if it preserves the Euler-Lagrangian operator \mathcal{E}_L . Let u be a projectable vector-field on $Y \rightarrow X$. The corresponding local isomorphisms of Y are generalized invariant transformations iff $\mathfrak{L}_{\bar{u}}L = h_0(\epsilon)$, where ϵ is a closed n -form on the bundle $Y \rightarrow X$. In this case, the weak transformation law (5.249) reads

$$s^*\epsilon \approx d(s^*u] \Xi_L)$$

for every critical section s of $Y \rightarrow X$. In particular, if $\epsilon = d\epsilon$ is an exact form, we get the *weak conservation law*

$$d(s^*(u)\lrcorner \Xi_L - \epsilon) \approx 0.$$

In particular, gauge transformations in gauge theory on a 3D base X are the invariant transformations if L is the *Yang–Mills Lagrangian density* and they are the generalized invariant transformations if L is the *Chern–Simons Lagrangian density*.

5.9.2 General Covariance Condition

Now we consider the class of bundles $T \rightarrow X$ which admit the canonical lift of vector-fields τ on X . They are called the bundles of geometrical objects. In fact, such canonical lift is the particular case of the horizontal lift of a field τ with respect to the suitable connection on the bundle $T \rightarrow X$ [Giachetta *et. al.* (2005)].

Let $\tau = \tau^\alpha \partial_\alpha$ be a vector-field on the manifold X . There exists the canonical lift

$$\tilde{\tau} = T\tau = \tau^\alpha \partial_\alpha + \partial_\nu \tau^\alpha \dot{x}^\nu \frac{\partial}{\partial \dot{x}^\alpha} \quad (5.253)$$

of τ onto the tangent bundle TX of X . This lift consists with the horizontal lift of τ by means the symmetric connection K on the tangent bundle which has τ as the integral section or as the geodesic field:

$$\partial_\nu \tau^\alpha + K^\alpha_{\alpha\nu} \tau^\alpha = 0.$$

Generalizing the canonical lift (5.253), one can construct the canonical lifts of a vector-field τ on X onto the following bundles over X . For the sake of simplicity, we denote all these lifts by the same symbol $\tilde{\tau}$. We have:

- the canonical lift of τ onto the cotangent bundle T^*X , given by

$$\tilde{\tau} = \tau^\alpha \partial_\alpha - \partial_\beta \tau^\nu \dot{x}_\nu \frac{\partial}{\partial \dot{x}_\beta};$$

- the canonical lift of τ onto the tensor bundle $T_m^k X = (\otimes^m TX) \otimes (\otimes^k T^*X)$, given by

$$\tilde{\tau} = \tau^\alpha \partial_\alpha + [\partial_\nu \tau^{\alpha_1} \dot{x}_{\beta_1 \dots \beta_k}^{\nu \alpha_2 \dots \alpha_m} + \dots - \partial_{\beta_1} \tau^\nu \dot{x}_{\nu \beta_2 \dots \beta_k}^{\alpha_1 \dots \alpha_m} - \dots] \frac{\partial}{\partial \dot{x}_{\beta_1 \dots \beta_k}^{\alpha_1 \dots \alpha_m}};$$

- the canonical lift of τ onto the bundle C of the linear connections on TX , given by

$$\tilde{\tau} = \tau^\alpha \partial_\alpha + [\partial_\nu \tau^\alpha k^\nu_{\beta\alpha} - \partial_\beta \tau^\nu k^\alpha_{\nu\alpha} - \partial_\alpha \tau^\nu k^\alpha_{\beta\nu} - \partial_{\beta\alpha} \tau^\alpha] \frac{\partial}{\partial k^\alpha_{\beta\alpha}}.$$

One can think of the vector-fields $\tilde{\tau}$ on a bundle of geometrical objects T as being the vector-fields associated with local 1-parameter groups of the holonomic isomorphisms of T induced by diffeomorphisms of its base X . In particular, if $T = TX$ they are the tangent isomorphisms. We call these isomorphisms the *general covariant transformations*.

Let T be the bundle of geometrical objects and L a Lagrangian density on the configuration space $J^1(X, T)$. Given a vector-field τ on the base X and its canonical lift $\tilde{\tau}$ onto T , one may use the first variational formula (5.247) in order to get the corresponding SEM transformation law. The left side of this formula can be simplified if the Lagrangian density satisfies the general covariance condition.

Note that, if the Lagrangian density L depends on background fields, we should consider the corresponding total bundle (5.238) and the Lagrangian density L_{tot} on the total configuration space $J^1(X, T)_{tot}$. We say that the Lagrangian density L satisfies the general covariance condition if L_{tot} is invariant under 1-parameter groups of general covariant transformations of T_{tot} induced by diffeomorphisms of the base X . It takes place iff, for any vector-field τ on X , the Lagrangian density L_{tot} obeys the equality

$$\mathfrak{L}_{j_0^1 \tilde{\tau}} L_{tot} = 0 \quad (5.254)$$

where $\tilde{\tau}$ is the canonical lift of τ onto T_{tot} and $j_0^1 \tilde{\tau}$ is the jet lift of $\tilde{\tau}$ onto $J^1(X, T)_{tot}$.

If the Lagrangian density L does not depend on background fields, the equality (5.254) becomes

$$\mathfrak{L}_{j_0^1 \tilde{\tau}} L = 0. \quad (5.255)$$

Substituting it in the first variational formula (5.247), we get the weak conservation law

$$d_H h_0(\tilde{\tau} | \Xi_L) \approx 0. \quad (5.256)$$

One can show that the conserved quantity is reduced to a superpotential term.

Here, we verify this fact in case of a tensor bundle $T \rightarrow X$. Let it be coordinated by (x^α, y^A) where the collective index A is employed. Given a

vector-field τ on X , its canonical lift $\tilde{\tau}$ on T reads

$$\tilde{\tau} = \tau^\alpha \partial_\alpha + u^{A\beta} \partial_\beta \tau^\alpha \partial_A.$$

Let a Lagrangian density L on the configuration space $J^1(X, T)$ be invariant under general covarian transformations. Then, it satisfies the equality (5.255) which takes the coordinate form

$$\partial_\alpha(\tau^\alpha \mathcal{L}) + u^{A\beta} \partial_\beta \tau^\alpha \partial_A \mathcal{L} + \hat{\partial}_\alpha(u^{A\beta} \partial_\beta \tau^\alpha) \partial_A^\alpha \mathcal{L} - y_\alpha^A \partial_\beta \tau^\alpha \partial_A^\beta \mathcal{L} = 0. \quad (5.257)$$

Due to the arbitrariness of the functions τ^α , the equality (5.257) is equivalent to the system of the equalities

$$\begin{aligned} \partial_\alpha \mathcal{L} &= 0, \\ \delta_\alpha^\beta \mathcal{L} + u^{A\beta} \partial_A \mathcal{L} + \hat{\partial}_\alpha(u^{A\beta} \partial_A^\alpha \mathcal{L} - y_\alpha^A \partial_A^\beta \mathcal{L}) &= 0, \end{aligned} \quad (5.258)$$

$$u^{A\beta} \partial_A^\alpha \mathcal{L} + u^{A\alpha} \partial_A^\beta \mathcal{L} = 0. \quad (5.259)$$

Note that the equality (5.258) can be brought into the form

$$\delta_\alpha^\beta \mathcal{L} + u^{A\beta} \delta_A \mathcal{L} + \hat{\partial}_\alpha(u^{A\beta} \partial_A^\alpha \mathcal{L}) = y_\alpha^A \partial_A^\beta \mathcal{L}, \quad (5.260)$$

where $\delta_A \mathcal{L}$ are the variational derivatives of the Lagrangian density L . Substituting the relations (5.260) and (5.259) into the weak identity

$$\hat{\partial}_\alpha[(u^{A\beta} \partial_\beta \tau^\alpha - y_\alpha^A \tau^\alpha) \partial_A^\alpha \mathcal{L} + \tau^\alpha \mathcal{L}] \approx 0,$$

we get the conservation law

$$\hat{\partial}_\alpha[-u^{A\alpha} \delta_A \mathcal{L} \tau^\alpha - \hat{\partial}_\alpha(u^{A\alpha} \partial_A^\alpha \mathcal{L} \tau^\alpha)] \approx 0, \quad (5.261)$$

where the conserved current is reduced to the superpotential term

$$Q_{\tilde{\tau}}^\alpha = -u^{A\alpha} \delta_A \mathcal{L} \tau^\alpha - \hat{\partial}_\alpha(u^{A\alpha} \partial_A^\alpha \mathcal{L} \tau^\alpha). \quad (5.262)$$

For general field models, we have the product $T \times Y$ of a bundle $T \rightarrow X$ of geometrical objects and some other bundle $Y \rightarrow X$. The lift of a vector-field τ on the base X onto the corresponding configuration space $J^1(X, T) \times J^1(X, Y)$ reads

$$\bar{\tau} = j_0^1 \tilde{\tau} + \tau^\alpha \Gamma_\alpha^i \partial_i + (\partial_\alpha(\tau^\alpha \Gamma_\alpha^i) + \tau^\alpha y_\alpha^j \partial_j \Gamma_\alpha^i - y_\alpha^i \partial_\alpha \tau^\alpha) \partial_i^\alpha$$

where Γ is a connection on the fibre bundle $Y \rightarrow X$.

In this case, we cannot say anything about the general covariance condition independently on the invariance of a Lagrangian density with respect to the internal symmetries.

On the other hand, in *gauge theory* (see subsection 5.11 below), several types of gauge transformations are considered. To get the Noether conservation laws, we restrict our consideration to vertical isomorphisms of the principal bundle P . These are the G -equivariant isomorphism Φ of P over Id_X , that is,

$$r_g \circ \Phi = \Phi \circ r_g, \quad (g \in G). \quad (5.263)$$

We call them the gauge isomorphisms. As is well-known, they yield the vertical isomorphisms of the bundle of principal connections C and the P -associated bundle E .

For example, let $P \rightarrow X$ be a principal bundle with the structure Lie group G . Let us consider general gauge isomorphisms Φ of this principal bundle over diffeomorphisms of the base X . They satisfy the relation (5.263). We denote by u_G the projectable vector-fields on P corresponding to local 1-parameter groups of such isomorphisms. There is the 1-1 correspondence between these vector-fields and sections of the bundle $T^G P = TP/G$. They are called the *general principal vector-fields* (see [Giachetta and Mangiarotti (1990)]). In particular, one can show that, given a vector-field τ on the base X , its horizontal lift onto the principal bundle P by means of a principal connection on P is a general principal vector-field.

General gauge isomorphisms of the principal bundle P , as like as its vertical isomorphisms, yield the corresponding isomorphisms of the associated bundles E and the bundle of principal connections C . We denote by the same symbol u_G the corresponding general principal vector-fields on these bundles.

Consider the product $S = C \times E \times T$, where $T \rightarrow X$ is a bundle of geometrical objects. Let a Lagrangian density L on the corresponding configuration space $J^1(T, S)$ be invariant under the isomorphisms of the bundle S which are general gauge isomorphisms of $C \times E$ over diffeomorphisms of the base X and the general covariant transformations of T induced by these diffeomorphisms of X . In particular, vertical isomorphisms of S consist of vertical isomorphisms of $C \times E$ only. It should be emphasized that the general gauge isomorphisms of the bundle $C \times E$ and those of the bundle T taken separately are not the bundle isomorphisms of the product S because they must covering the same diffeomorphisms of the base X of Y . At the same time, one can say that the Lagrangian density L satisfies the general covariance condition in the sense that it is invariant under general isomorphisms of the bundle S [Giachetta and Mangiarotti (1990)].

This is phrased in terms of the Lie derivatives as follows. Let

$$u_G = \tau^\alpha \partial_\alpha + u^A \partial_A$$

be a general principal vector-field on the product $C \times E$ which is projected onto the vector-field $\tau = \tau^\alpha \partial_\alpha$ on the base X . The corresponding general principal vector-field on the bundle Y reads

$$\tilde{u}_G = \tilde{\tau} + u^A \partial_A, \quad (5.264)$$

where $\tilde{\tau}$ is the canonical lift of τ onto the bundle of geometrical objects T . A Lagrangian density L is invariant under general isomorphisms of the bundle S iff

$$\mathfrak{L}_{j_0^1 \tilde{u}_G} L = 0, \quad (5.265)$$

where the jet lift $j_0^1 \tilde{u}_G$ of the vector-field \tilde{u}_G takes the coordinate form

$$j_0^1 \tilde{u}_G = j_0^1 \tilde{\tau} - y_\alpha^A \partial_\alpha u^A \partial_A^\alpha + u^A \partial_A + \hat{\partial}_\alpha u^A \partial_A^\alpha.$$

There are the topological field theories, besides the gravitation theory, where we can use the condition (5.265) (see subsection 5.11.8 below).

5.10 Application: Jets and Hamiltonian Field Theory

Recall that the Hamiltonian counterpart of the classical Lagrangian field theory (see subsection 5.9 above) is the *covariant Hamiltonian field theory*, in which momenta correspond to derivatives of fields with respect to all world coordinates. It is well-known that classical Lagrangian and covariant Hamiltonian field theories are equivalent in the case of a hyperregular Lagrangian, and they are quasi-equivalent if a Lagrangian is almost-regular (see [Sardanashvily (1993); Sardanashvily (1995); Giachetta *et. al.* (1997); Giachetta *et. al.* (1999); Mangiarotti and Sardanashvily (2000a); Sardanashvily (2002a)]). Further, in order to quantize covariant Hamiltonian field theory, one usually attempts to construct and quantize a multisymplectic generalization of the Poisson bracket. The path-integral quantization of covariant Hamiltonian field theory was recently suggested in [Bashkirov and Sardanashvily (2004)].

Recall that the symplectic Hamiltonian technique applied to field theory leads to instantaneous Hamiltonian formalism on an infinite-dimensional phase-space coordinated by field functions at some instant of time (see [Gotay (1991a)] for the strict mathematical exposition of this formalism). The

true Hamiltonian counterpart of classical first-order Lagrangian field theory is covariant Hamiltonian formalism, where canonical momenta p_i^μ correspond to derivatives y_μ^i of fields y^i with respect to all world coordinates x^μ . This formalism has been developed since the 1970s in its polysymplectic, multisymplectic and Hamilton–de Donder variants (see [Giachetta *et. al.* (1997); Lopez and Marsden (2003)]). In order to quantize covariant Hamiltonian field theory, one usually attempts to construct multisymplectic generalization of the Poisson bracket with respect to the derivatives $\partial/\partial y^i$ and $\partial/\partial p_i^\mu$ [Kanatchikov (1998)].

We can also quantize covariant Hamiltonian field theory in path–integral terms following [Bashkirov and Sardanashvily (2004)]. A polysymplectic Hamiltonian system with a Hamiltonian $\mathcal{H}(x^\mu, y^i, p_i^\mu)$ is equivalent to a Lagrangian system with the Lagrangian

$$\mathcal{L}_{\mathcal{H}}(x^\mu, y^i, p_i^\mu, y_\alpha^i) = p_i^\alpha y_\alpha^i - \mathcal{H}(x^\mu, y^i, p_i^\mu, y_\alpha^i) \quad (5.266)$$

of the variables y^i and p_i^μ . In subsection 6.3.10 below we will quantize this Lagrangian system in the framework of *perturbative quantum field theory*. Briefly, if there is no constraint and the matrix $\partial^2 \mathcal{H}/\partial p_i^\mu \partial p_j^\nu$ is nondegenerate and positive–definite, this quantization is given by the generating functional

$$Z = \mathcal{N}^{-1} \int \exp\left\{ \int (\mathcal{L}_{\mathcal{H}} + \Lambda + iJ_i y^i + iJ_\mu^\mu p_i^\mu) dx \right\} \prod_x [dp(x)][dy(x)] \quad (5.267)$$

of Euclidean Green functions, where Λ comes from the normalization condition

$$\int \exp\left\{ \int \left(-\frac{1}{2} \partial_\mu^i \partial_\nu^j \mathcal{H} p_i^\mu p_j^\nu + \Lambda\right) dx \right\} \prod_x [dp(x)] = 1.$$

If a Hamiltonian \mathcal{H} is degenerate, the Lagrangian $\mathcal{L}_{\mathcal{H}}$ (5.266) may admit gauge symmetries. In this case, integration of a generating functional along gauge group orbits must be finite. If there are constraints, the Lagrangian system with a Lagrangian $\mathcal{L}_{\mathcal{H}}$ (5.266) restricted to the constraint manifold is quantized.

In order to verify this path–integral quantization scheme, we apply it to Hamiltonian field systems associated to Lagrangian field systems with quadratic Lagrangians

$$\mathcal{L} = \frac{1}{2} a_{ij}^{\lambda\mu} y_\alpha^i y_\mu^j + b_i^\alpha y_\alpha^i + c, \quad (5.268)$$

where a , b and c are functions of world coordinates x^μ and field variables y^i . Note that, in the framework of perturbative quantum field theory, any Lagrangian is split into the sum of a quadratic Lagrangian (5.268) and an interaction term quantized as a perturbation.

For example, let the Lagrangian (5.268) be hyperregular, i.e., the matrix function a is nondegenerate. Then there exists a unique associated Hamiltonian system whose Hamiltonian \mathcal{H} is quadratic in momenta p_i^μ , and so is the Lagrangian $\mathcal{L}_{\mathcal{H}}$ (5.266). If the matrix function a is positive-definite on an Euclidean space-time, the generating functional (5.267) is a Gaussian integral of momenta $p_i^\mu(x)$. Integrating Z with respect to $p_i^\mu(x)$, one restarts the generating functional of quantum field theory with the original Lagrangian \mathcal{L} (5.268). We extend this result to field theories with almost-regular Lagrangians \mathcal{L} (5.268), e.g., *Yang–Mills gauge theory*. The key point is that, though such a Lagrangian \mathcal{L} induces constraints and admits different associated Hamiltonians \mathcal{H} , all the Lagrangians $\mathcal{L}_{\mathcal{H}}$ coincide on the constraint manifold, and we have a unique constrained Hamiltonian system which is quasi-equivalent to the original Lagrangian one [Giachetta *et. al.* (1997)].

5.10.1 Covariant Hamiltonian Field Systems

To develop the covariant Hamiltonian field theory suitable for path-integral quantization, we start by following the geometrical formulation of classical field theory (see [Sardanashvily (1993); Sardanashvily (1995); Giachetta *et. al.* (1997); Mangiarotti and Sardanashvily (2000a); Sardanashvily (2002a)]), in which classical fields are represented by sections of fibre bundles. Let $Y \rightarrow X$ be a smooth fibre bundle provided with bundle coordinates (x^μ, y^i) . Recall from subsection 5.9 above, that the configuration space of Lagrangian field theory on Y is the 1-jet space $J^1(X, Y)$ of Y . It is equipped with the bundle coordinates (x^μ, y^i, y_μ^i) compatible with the composite fibration

$$J^1(X, Y) \xrightarrow{\pi_0^1} Y \xrightarrow{\pi} X.$$

Any section $s : X \rightarrow Y$ of a fibre bundle $Y \rightarrow X$ is prolonged to the section $j^1s : X \rightarrow J^1(X, Y)$ of the jet bundle $J^1(X, Y) \rightarrow X$, such that $y_\mu^i \circ j^1s = \partial_\mu s^i$.

Also, recall that the first-order Lagrangian is defined as a *horizontal*

density

$$L = \mathcal{L}\omega : J^1(X, Y) \rightarrow \wedge^n T^*X, \quad (\omega = dx^1 \wedge \cdots \wedge dx^n, n = \dim X), \quad (5.269)$$

on the jet space $J^1(X, Y)$. The corresponding *Euler–Lagrangian equations*

$$(\partial_i - d_\alpha \partial_i^\alpha) \mathcal{L} = 0, \quad d_\alpha = \partial_\alpha + y_\alpha^i \partial_i + y_{\lambda\mu}^i \partial_i^\mu, \quad (5.270)$$

represent the subset of the 2-jet space $J^2(X, Y)$ of Y , coordinated by $(x^\mu, y^i, y_\alpha^i, y_{\lambda\mu}^i)$. A section s of $Y \rightarrow X$ is a solution of these equations if its second jet prolongation $j^2 s$ lives in the subset (5.270).

The phase-space of covariant (polysymplectic) Hamiltonian field theory on a fibre bundle $Y \rightarrow X$ is the *Legendre bundle* (see (5.225) above)

$$\Pi = \wedge^n T^*X \otimes V^*Y \otimes TX = V^*Y \wedge (\wedge^{n-1} T^*X), \quad (5.271)$$

where V^*Y is the vertical cotangent bundle of $Y \rightarrow X$. The Legendre bundle Π is equipped with the holonomic bundle coordinates (x^α, y^i, p_i^μ) compatible with the *composite fibration*

$$\Pi \xrightarrow{\pi_Y} Y \xrightarrow{\pi} X, \quad (5.272)$$

admitting the canonical polysymplectic form

$$\Omega = dp_i^\alpha \wedge dy^i \wedge \omega \otimes \partial_\alpha.$$

A *covariant Hamiltonian* \mathcal{H} on Π (5.272) is defined as a section $p = -\mathcal{H}$ of the trivial line bundle (i.e., 1D fibre bundle)

$$Z_Y = T^*Y \wedge (\wedge^{n-1} T^*X) \rightarrow \Pi, \quad (5.273)$$

equipped with holonomic bundle coordinates $(x^\alpha, y^i, p_i^\mu, p)$. This fibre bundle admits the canonical multisymplectic Liouville form

$$\Xi = p\omega + p_i^\alpha dy^i \wedge \omega_\alpha, \quad \text{with} \quad \omega_\alpha = \partial_\alpha \rfloor \omega.$$

The pull-back of Ξ onto Π by a Hamiltonian \mathcal{H} gives the *Hamiltonian form*

$$H = \mathcal{H}^* \Xi_Y = p_i^\alpha dy^i \wedge \omega_\alpha - \mathcal{H}\omega \quad (5.274)$$

on Π . The corresponding *covariant Hamiltonian equations* on Π ,

$$y_\alpha^i = \partial_\alpha^i \mathcal{H}, \quad p_{\lambda i}^\alpha = -\partial_i \mathcal{H}, \quad (5.275)$$

represent the closed submanifold of the jet space $J^1(X, \Pi)$ of Π . A section r of $\Pi \rightarrow X$ is a solution of these equations if its jet prolongation j^1r lives in the submanifold (5.275).

A section r of $\Pi \rightarrow X$ is a solution of the covariant Hamiltonian equations (5.275) iff it satisfies the condition $r^*(u \rfloor dH) = 0$ for any vertical vector-field u on $\Pi \rightarrow X$.

Alternatively, a section r of $\Pi \rightarrow X$ is a solution of the covariant Hamiltonian equations (5.275) iff it is a solution of the Euler–Lagrangian equations for the first-order Lagrangian $L_{\mathcal{H}}$ on $J^1(X, \Pi)$,

$$L_{\mathcal{H}} = h_0(H) = \mathcal{L}_{\mathcal{H}}\omega = (p_i^\alpha y_\alpha^i - \mathcal{H})\omega, \quad (5.276)$$

where h_0 sends exterior forms on Π onto horizontal exterior forms on $J^1(X, \Pi) \rightarrow X$, using the rule $h_0(dy^i) = y_\alpha^i dx^\alpha$.

Note that, for any section r of $\Pi \rightarrow X$, the pull-backs r^*H and $j^1r^*L_{\mathcal{H}}$ coincide. This fact motivated [Bashkirov and Sardanashevily (2004)] to quantize covariant Hamiltonian field theory with a Hamiltonian \mathcal{H} on Π as a Lagrangian system with the Lagrangian $L_{\mathcal{H}}$ (5.276).

Furthermore, let $i_N : N \rightarrow \Pi$ be a closed imbedded subbundle of the Legendre bundle $\Pi \rightarrow Y$ which is regarded as a constraint space of a covariant Hamiltonian field system with a Hamiltonian \mathcal{H} . This Hamiltonian system is restricted to N as follows. Let $H_N = i_N^*H$ be the pull-back of the Hamiltonian form H (5.274) onto N . The constrained Hamiltonian form H_N defines the constrained Lagrangian

$$L_N = h_0(H_N) = (j^1i_N)^*L_{\mathcal{H}} \quad (5.277)$$

on the jet space $J^1(X, N_L)$ of the fibre bundle $N_L \rightarrow X$. The Euler–Lagrangian equations for this Lagrangian are called the *constrained Hamiltonian equations*.

Note that, the Lagrangian $L_{\mathcal{H}}$ (5.276) is the pull-back onto $J^1(X, \Pi)$ of the horizontal form $L_{\mathcal{H}}$ on the bundle product $\Pi \times J^1(X, Y)$ over Y by the canonical map $J^1(X, \Pi) \rightarrow \Pi \times J^1(X, Y)$. Therefore, the constrained Lagrangian L_N (5.277) is the restriction of $L_{\mathcal{H}}$ to $N \times J^1(X, Y)$.

A section r of the fibre bundle $N \rightarrow X$ is a solution of constrained Hamiltonian equations iff it satisfies the condition $r^*(u_N \rfloor dH) = 0$ for any vertical vector-field u_N on $N \rightarrow X$.

Any solution of the covariant Hamiltonian equations (5.275) which lives in the constraint manifold N is also a solution of the constrained Hamiltonian equations on N . This fact motivates us to quantize covariant Hamil-

tonian field theory on a constraint manifold N as a Lagrangian system with the pull-back Lagrangian L_N (5.277).

Since a constraint manifold is assumed to be a closed imbedded submanifold of Π , there exists its open neighborhood U which is a fibre bundle $U \rightarrow N$. If Π is a fibre bundle $\pi_N : \Pi \rightarrow N$ over N , it is often convenient to quantize a Lagrangian system on Π with the pull-back Lagrangian $\pi_N^* L_N$, but integration of the corresponding generating functional along the fibres of $\Pi \rightarrow N$ must be finite.

In order to verify this quantization scheme, let us associate to a Lagrangian field system on Y a covariant Hamiltonian system on Π , then let us quantize this Hamiltonian system and compare this quantization with that of an original Lagrangian system.

5.10.2 Associated Lagrangian and Hamiltonian Systems

In order to relate classical Lagrangian and covariant Hamiltonian field theories, let us recall that, besides the Euler–Lagrangian equations, a Lagrangian L (5.269) also induces the Cartan equations which are given by the subset

$$\begin{aligned} (\bar{y}_\mu^j - y_\mu^j) \partial_i^\alpha \partial_j^\mu \mathcal{L} &= 0, \\ \partial_i \mathcal{L} - \bar{d}_\alpha \partial_i^\alpha \mathcal{L} + (\bar{y}_\mu^j - y_\mu^j) \partial_j^\mu \mathcal{L} &= 0, \quad \bar{d}_\alpha = \partial_\alpha + \bar{y}_\alpha^i \partial_i + \bar{y}_{\lambda\mu}^i \partial_i^\mu, \end{aligned} \quad (5.278)$$

of the repeated jet space $J^1 J^1(X, Y)$ coordinated by $(x^\mu, y^i, y_\lambda^i, \bar{y}_\alpha^i, \bar{y}_{\lambda\mu}^i)$. A solution of the Cartan equations is a section \bar{s} of the jet bundle $J^1(X, Y) \rightarrow X$ whose jet prolongation $j^1 \bar{s}$ lives in the subset (5.278). Every solution s of the Euler–Lagrangian equations (5.270) defines the solution $j^1 s$ of the Cartan equations (5.278). If \bar{s} is a solution of the Cartan equations and $\bar{s} = j^1 s$, then s is a solution of the Euler–Lagrangian equations. If a Lagrangian L is regular, the equations (5.270) and (5.278) are equivalent.

Recall that any Lagrangian L (5.269) induces the Legendre map (5.226), i.e.,

$$\widehat{L} : J^1(X, Y) \longrightarrow \Pi, \quad p_i^\alpha \circ \widehat{L} = \partial_i^\alpha \mathcal{L}, \quad (5.279)$$

over Id_Y whose image $N_L = \widehat{L}(J^1(X, Y))$ is called the Lagrangian constraint space. A Lagrangian L is said to be hyperregular if the Legendre map (5.279) is a diffeomorphism. A Lagrangian L is called almost-regular if the Lagrangian constraint space is a closed imbedded subbundle $i_N : N_L \rightarrow \Pi$ of the Legendre bundle $\Pi \rightarrow Y$ and the surjection

$\widehat{L} : J^1(X, Y) \rightarrow N_L$ is a submersion (i.e., a fibre bundle) whose fibres are connected. Conversely, any Hamiltonian \mathcal{H} induces the Hamiltonian map

$$\widehat{H} : \Pi \rightarrow J^1(X, Y), \quad y_\alpha^i \circ \widehat{H} = \partial_\alpha^i \mathcal{H}. \quad (5.280)$$

A Hamiltonian \mathcal{H} on Π is said to be associated to a Lagrangian L on $J^1(X, Y)$ if \mathcal{H} satisfies the following relations (with $(x^\mu, y^i, p_i^\mu) \in N_L$)

$$\widehat{L} \circ \widehat{H} \circ \widehat{L} = \widehat{L}, \quad p_i^\mu = \partial_i^\mu \mathcal{L}(x^\mu, y^i, \partial_\alpha^j \mathcal{H}), \quad (5.281)$$

$$\widehat{H}^* L_{\mathcal{H}} = \widehat{H}^* L, \quad p_i^\mu \partial_\mu^i \mathcal{H} - \mathcal{H} = \mathcal{L}(x^\mu, y^j, \partial_\alpha^j \mathcal{H}). \quad (5.282)$$

If an associated Hamiltonian \mathcal{H} exists, the Lagrangian constraint space N_L is given by the coordinate relations (5.281) and $\widehat{H} \circ \widehat{L}$ is a projector from Π onto N_L .

For example, any hyperregular Lagrangian L admits a unique associated Hamiltonian \mathcal{H} such that

$$\widehat{H} = \widehat{L}^{-1}, \quad \mathcal{H} = p_i^\mu \widehat{L}_\mu^i{}^{-1} - \mathcal{L}(x^\alpha, y^i, \widehat{L}_\alpha^i{}^{-1}).$$

In this case, any solution s of the Euler–Lagrangian equations (5.270) defines the solution $r = \widehat{L} \circ j^1 s$, of the covariant Hamiltonian equations (5.275). Conversely, any solution r of these Hamiltonian equations induces the solution $s = \pi_Y \circ r$ of the Euler–Lagrangian equations (5.270).

A degenerate Lagrangian need not admit an associated Hamiltonian. If such a Hamiltonian exists, it is not necessarily unique. Let us restrict our consideration to almost-regular Lagrangians. From the physical viewpoint, the most of Lagrangian field theories is of this type. From the mathematical one, this notion of degeneracy is particularly appropriate for the study of relations between Lagrangian and covariant Hamiltonian formalisms as follows [Bashkirov and Sardanashvily (2004)].

Let L be an almost-regular Lagrangian and \mathcal{H} an associated Hamiltonian. Let a section r of $\Pi \rightarrow X$ be a solution of the covariant Hamiltonian equations (5.275) for \mathcal{H} . If r lives in the constraint manifold N_L , then $s = \pi_Y \circ r$ satisfies the Euler–Lagrangian equations (5.270) for L , while $\bar{s} = \widehat{H} \circ r$ obeys the Cartan equations (5.278). Conversely, let \bar{s} be a solution of the Cartan equations (5.278) for L . If \mathcal{H} satisfies the relation

$$\widehat{H} \circ \widehat{L} \circ \bar{s} = j^1(\pi_0^1 \circ \bar{s}),$$

the section $r = \widehat{L} \circ \bar{s}$ of the Legendre bundle $\Pi \rightarrow X$ is a solution of the Hamiltonian equations (5.275) for \mathcal{H} . If $\bar{s} = j^1 s$, we get the relation between

solutions the Euler–Lagrangian equations and the covariant Hamiltonian ones.

Due to this Theorem, one needs a set of different associated Hamiltonians in order to recover all solutions of the Euler–Lagrangian and Cartan equations for an almost-regular Lagrangian L . We can overcome this ambiguity as follows.

Let $\mathcal{H}, \mathcal{H}'$ be two different Hamiltonians associated to an almost-regular Lagrangian L . Let H, H' be the corresponding Hamiltonian forms (5.274). Their pull-backs i_N^*H and i_N^*H' onto the Lagrangian constraint manifold N_L coincide with each other.

It follows that, if an almost-regular Lagrangian admits associated Hamiltonians \mathcal{H} , it defines a unique constrained Hamiltonian form $H_N = i_N^*H$ on the Lagrangian constraint manifold N_L and a unique constrained Lagrangian $L_N = h_0(H_N)$ (5.277) on the jet space $J^1(X, N_L)$ of the fibre bundle $N_L \rightarrow X$. For any Hamiltonian \mathcal{H} associated to L , every solution r of the Hamiltonian equations which lives in the Lagrangian constraint space N_L is a solution of the constrained Hamiltonian equations for L_N .

Let an almost-regular Lagrangian L admit associated Hamiltonians. A section \bar{s} of the jet bundle $J^1(X, Y) \rightarrow X$ is a solution of the Cartan equations for L iff $\widehat{L} \circ \bar{s}$ is a solution of the constrained Hamiltonian equations. In particular, any solution r of the constrained Hamiltonian equations gives the solution $\bar{s} = \widehat{H} \circ r$ of the Cartan equations. This Theorem shows that the constrained Hamiltonian equations and the Cartan equations are quasi-equivalent. Thus, one can associate to an almost-regular Lagrangian L (5.269) a unique constrained Lagrangian system on the constraint Lagrangian manifold N_L . Let us compare quantizations of these Lagrangian systems on Y and $N_L \subset \Pi$ in the case of an almost-regular quadratic Lagrangian L .

5.10.3 Evolution Operator

Recall that the covariant Hamiltonian field theory is mainly developed in its multisymplectic and polysymplectic variants, related to the two different Legendre maps in the first-order calculus of variations on fibre bundles [Sardanashvily (2002c)] (also, see [Gotay (1991a)] for a survey of symplectic formalism).

Recall that, given a fibre bundle $Y \rightarrow X$ coordinated by (x^α, y^i) , every first-order Lagrangian $L : J^1(X, Y) \rightarrow \wedge^n T^*X$ is given by $L = \mathcal{L}\omega$, $\omega = dx^1 \wedge \cdots \wedge dx^n$, ($n = \dim X$), and induces the Legendre map of the 1-jet

space $J^1(X, Y)$ to the Legendre bundle

$$\Pi = \wedge^n T^*X \otimes V^*Y \otimes TX, \quad (5.283)$$

coordinated by $(x^\alpha, y^i, p_i^\alpha)$. The Π admits the *canonical polysymplectic form*

$$\Omega_\Pi = dp_i^\alpha \wedge dy^i \wedge \omega \otimes \partial_\alpha, \quad (5.284)$$

and is regarded as the polysymplectic phase-space of fields.

The multisymplectic phase-space of fields is the homogeneous Legendre bundle

$$Z = T^*Y \wedge (\wedge^{n-1} T^*X), \quad (5.285)$$

coordinated by $(x^\alpha, y^i, p_i^\alpha, p)$ and equipped with the canonical multisymplectic form

$$\Omega = d\Xi = dp \wedge \omega + dp_i^\alpha \wedge dy^i \wedge \omega_\alpha, \quad \text{with} \quad \omega_\alpha = \partial_\alpha \rfloor \omega. \quad (5.286)$$

It is natural that one attempts to generalize a Poisson bracket on symplectic manifolds to polysymplectic and multisymplectic manifolds in order to get the covariant canonical quantization of field theory. Different variants of such a bracket have been suggested. However, it seems that no canonical bracket corresponds to the TX -valued polysymplectic form (5.284), unless $\dim X = 1$. On the contrary, using the exterior multisymplectic form (5.286), one can associate multivector-fields to exterior forms on the fibre bundle Z (6.212) (but not to all of them), and can introduce a desired bracket of these forms via the Schouten–Nijenhuis bracket of multivector-fields.

Note that no bracket determines the evolution operator in polysymplectic and multisymplectic Hamiltonian formalism, including the case of $\dim X = 1$ of the time-dependent mechanics (see subsection 5.6.1 above). Written as a bracket, the evolution operator can be quantized, and it determines the *Heisenberg equation*.³

³Recall that in the (Lorentz-invariant) *Heisenberg quantum picture*, the state vector $|\psi\rangle$ does not change with time, and an observable A satisfies the Heisenberg equation of motion

$$\dot{A} = (i\hbar)^{-1}[A, H] + (\partial_t A)_{\text{classical}},$$

which becomes the classical Poisson equation if we replace its commutator $[A, H]$ by the Poisson bracket A, H .

Recall the following relationship between first-order dynamical equations, connections, multivector-fields and evolution operators on a fibre bundle.

(i) Let $\pi : Q \rightarrow X$ be a fibre bundle coordinated by (x^μ, q^a) . As a section $\gamma : Q \rightarrow J^1(X, Q)$ of the 1-jet bundle $J^1(X, Q) \rightarrow Q$, any connection

$$\gamma = dx^\mu \otimes (\partial_\mu + \gamma_\mu^a \partial_a), \quad (5.287)$$

on $Q \rightarrow X$ defines the first-order differential operator

$$D : J^1(X, Q) \rightarrow T^*X \otimes VQ, \quad (x^\mu, q^a, q_\mu^a) \rightarrow (x^\mu, q^a, q_\mu^a - \gamma_\mu^a(x^\nu, q^b)) \quad (5.288)$$

on $Q \rightarrow X$ called the *covariant differential* with respect to γ . The kernel of this differential operator is a closed imbedded subbundle of $J^1(X, Q) \rightarrow X$, given by the first-order dynamical equation

$$q_\mu^a - \gamma_\mu^a(x^\nu, q^b) = 0 \quad (5.289)$$

on a fibre bundle $Q \rightarrow X$. Conversely, any first-order dynamical equation on $Q \rightarrow X$ is of this type.

(ii) Let $HQ \subset TQ$ be the horizontal distribution defined by a connection γ . If X is orientable, there exists a nowhere vanishing global section of the exterior product $\wedge^n HQ \rightarrow Q$. It is a locally decomposable π -transverse n -vector-field on Q . Conversely, every multivector-field of this type on $Q \rightarrow X$ induces a connection and, consequently, a first-order dynamical equation on this fibre bundle [Echeverría *et. al.* (1998)].

(iii) Given a first-order dynamical equation γ on a fibre bundle $Q \rightarrow X$, the corresponding *evolution operator* d_γ is defined as the pull-back d_γ onto the shell of the horizontal differential

$$d_H = dx^\mu (\partial_\mu + q_\mu^a \partial_a)$$

acting on smooth real functions on Q . It reads

$$d_\gamma f = (\partial_\mu + \gamma_\mu^a \partial_a) f dx^\mu, \quad (f \in C^\infty(Q)). \quad (5.290)$$

This expression shows that d_γ is projected onto Q , and it is a first-order differential operator on functions on Q . In particular, if a function f obeys the evolution equation $d_\gamma f = 0$, it is constant on any solution of the dynamical equation (5.289).

In Hamiltonian dynamics on Q , a problem is to represent the evolution operator (5.290) as a bracket of f with some exterior form on Q .

First of all, let us study the case of fibre bundles $Y \rightarrow X$ over a 1D orientable connected manifold X (i.e., X is either \mathbb{R} or S^1). In this case, the Legendre bundle Π (5.283) is isomorphic to the vertical cotangent bundle V^*Y of $Y \rightarrow X$ coordinated by (x, y^i, p_i) , and the polysymplectic form Ω_Π (5.284) on V^*Y reads

$$\Omega_\Pi = dp_i \wedge dy^i \wedge dx \otimes \partial_x. \quad (5.291)$$

Therefore, the homogeneous Legendre bundle (6.212) is the cotangent bundle T^*Y , coordinated by (x, y^i, p, p_i) , and the multisymplectic form (5.286) becomes the canonical symplectic form on T^*Y , given by

$$\Omega = dp \wedge dx + dp_i \wedge dy^i. \quad (5.292)$$

The vertical cotangent bundle V^*Y admits the canonical *Poisson bracket*

$$\{f, f'\}_V = \partial^i f \partial_i f' - \partial_i f \partial^i f', \quad (f, f' \in C^\infty(V^*Y)), \quad (5.293)$$

given by the relation

$$\zeta^* \{f, f'\}_V = \{\zeta^* f, \zeta^* f'\},$$

where $\{, \}$ is the canonical Poisson bracket on T^*Y [Mangiarotti and Sardanashvily (1998); Sardanashvily (1998)]. However, the Poisson structure (5.293) fails to determine Hamiltonian dynamics on the fibre bundle $V^*Q \rightarrow X$ because all Hamiltonian vector-fields with respect to this structure are vertical. At the same time, in accordance with general polysymplectic formalism [Giachetta *et. al.* (1997)], a section $h, p \circ h = -\mathcal{H}$, of the fibre bundle $V^*Y \rightarrow T^*Y$ induces a polysymplectic Hamiltonian form on V^*Y ,

$$H = p_i dy^i - \mathcal{H} dx. \quad (5.294)$$

The associated Hamiltonian connection on $V^*Y \rightarrow X$ with respect to the polysymplectic form (5.291) is

$$\gamma_H = dx \otimes (\partial_x + \partial^i \mathcal{H} \partial_i - \partial_i \mathcal{H} \partial^i). \quad (5.295)$$

It defines the Hamiltonian equation on V^*Y ,

$$y_x^i = \partial^i \mathcal{H}, \quad p_{ix} = -\partial_i \mathcal{H}.$$

The corresponding evolution operator (5.290) takes the local form

$$d_\gamma f = (\partial_x f + \{\mathcal{H}, f\}_V) dx, \quad (f \in C^\infty(V^*Q)). \quad (5.296)$$

The bracket $\{\mathcal{H}, f\}_V$ in this expression is not globally defined because \mathcal{H} is not a function on V^*Q . Therefore, the evolution operator (5.296) does not reduce to the Poisson bracket (5.293).

Let us now consider the pull-back ζ^*H of the Hamiltonian form H (5.294) onto T^*Y . Then the difference

$$H^* = \Xi - \zeta^*H = (p + \mathcal{H})dx \quad (5.297)$$

is a horizontal density on the fibre bundle $T^*Y \rightarrow X$. It is a multisymplectic Hamiltonian form. The corresponding Hamiltonian connection γ on $T^*Y \rightarrow X$ is given by the condition

$$\gamma(\Omega) = dH^*, \quad (5.298)$$

where the map

$$\gamma(\Omega) = dx \wedge [(\partial_x + \gamma^p \partial_p + \gamma^i \partial_i + \gamma_i \partial^i)]\Omega$$

is induced by an endomorphism of T^*Y determined by the tangent-valued form γ . We get

$$\gamma = dx \otimes (\partial_x + \gamma^p \partial_p + \partial^i \mathcal{H} \partial_i - \partial_i \mathcal{H} \partial^i), \quad (5.299)$$

where the coefficient γ^p is arbitrary. Note that this connection projects to the connection γ_H (5.295) on $V^*Y \rightarrow X$. As a consequence, it defines the evolution operator whose restriction to the pull-back of functions on V^*Q is exactly the evolution operator (5.296). But now this operator locally reduces to the Poisson bracket on T^*Y ,

$$d_\gamma f = \{p + \mathcal{H}, f\}dx, \quad (f \in C^\infty(V^*Y)). \quad (5.300)$$

However, this bracket is not globally defined, too, since $p + \mathcal{H}$ is a horizontal density, but not a function on T^*Y .

Let us introduce the function $\mathcal{E} = \rho^{-1}(p + \mathcal{H})$ on T^*Y , where ρdx is some nowhere vanishing density on X . The Hamiltonian vector-field of \mathcal{E} with respect to the symplectic form Ω on T^*Y reads

$$\vartheta_{\mathcal{E}} = \rho^{-1} \partial_x - \partial_x \mathcal{E} \partial^p + \partial^i \mathcal{E} \partial_i - \partial_i \mathcal{E} \partial^i.$$

This vector-field is horizontal with respect to the connection (5.299), where $\gamma_p = -\rho \partial_x \mathcal{E}$, and it determines this connection in the form

$$\gamma = dx \otimes (\partial_x - \rho \partial_x \mathcal{E} \partial^p + \rho \partial^i \mathcal{E} \partial_i - \rho \partial_i \mathcal{E} \partial^i).$$

Therefore, the evolution operator (5.300) is rewritten as

$$d_\gamma f = \rho \{ \mathcal{E}, f \} dx, \quad (5.301)$$

The bracket $\{ \mathcal{E}, f \}$ is well-defined, but d_γ does not equal this bracket because of the factor ρ .

The multisymplectic bracket with the horizontal density H^* (5.297) also cannot help us since there is no Hamiltonian multivector-field associated to H^* relative to the symplectic form Ω .

The manifolds $X = \mathbb{R}$ and $X = S^1$ can be equipped with coordinates x possessing transition functions $x' = x + \text{const}$, and one can always choose the density $\rho = 1$. Then the evolution operator (5.301) reduces to a Poisson bracket in full. If $X = \mathbb{R}$, this is the case of time-dependent mechanics where time reparametrization is forbidden [Mangiarotti and Sardanashvily (2000b); Giachetta *et. al.* (2002a)].

Now we turn to the general case of $\dim X > 1$. In the framework of polysymplectic formalism [Giachetta *et. al.* (1997)], a polysymplectic Hamiltonian form on the Legendre bundle Π (5.283) reads

$$H = p_i^\alpha dy^i \wedge \omega_\alpha - \mathcal{H}\omega. \quad (5.302)$$

The associated Hamiltonian connection

$$\gamma_H = dx^\alpha \otimes (\partial_\alpha + \gamma_\alpha^i \partial_i + \gamma_{i\lambda}^\mu \partial_\mu^i)$$

fails to be uniquely determined, but obeys the equations

$$\gamma_\alpha^i = \partial_\alpha^i \mathcal{H}, \quad \gamma_{i\lambda}^\alpha = -\partial_i \mathcal{H}.$$

The values of these connections assemble into a closed imbedded subbubble

$$y_\alpha^i = \partial_\alpha^i \mathcal{H}, \quad p_{i\lambda}^\alpha = -\partial_i \mathcal{H}$$

of the jet bundle $J^1(X, \Pi) \rightarrow X$ which is the first-order polysymplectic Hamiltonian equation on Π . This equation is not algebraically solved for the highest order derivatives and, therefore, it is not a dynamical equation. As a consequence, the evolution operator depends on the jet coordinates $p_{i\mu}^\alpha$ and, therefore, it is not a differential operator on functions on Π . Clearly, no bracket on Π can determine such an operator.

5.10.4 Quadratic Degenerate Systems

Given a fibre bundle $Y \rightarrow X$, let us consider a quadratic Lagrangian L (5.268), where a , b and c are local functions on Y . This property is coordinate-independent since $J^1(X, Y) \rightarrow Y$ is an affine bundle modelled over the vector bundle $T^*X \otimes VY$, where VY denotes the vertical tangent bundle of $Y \rightarrow X$. The associated Legendre map (5.279) reads

$$p_i^\alpha \circ \widehat{L} = a_{ij}^{\alpha\mu} y_\mu^j + b_i^\alpha. \quad (5.303)$$

Let a Lagrangian L (5.268) be almost-regular, i.e., the matrix function a is a linear bundle map

$$a : T^*X \otimes VY \rightarrow \Pi, \quad p_i^\alpha = a_{ij}^{\alpha\mu} \bar{y}_\mu^j, \quad (5.304)$$

of constant rank, where $(x^\alpha, y^i, \bar{y}_\alpha^i)$ are bundle coordinates on $T^*X \otimes VY$. Then the Lagrangian constraint space N_L (5.303) is an affine subbundle of the Legendre bundle $\Pi \rightarrow Y$ (5.283). Hence, $N_L \rightarrow Y$ has a global section. For the sake of simplicity, let us assume that it is the canonical zero section $\widehat{0}(Y)$ of $\Pi \rightarrow Y$. The kernel of the Legendre map (5.303) is also an affine subbundle of the affine jet bundle $J^1(X, Y) \rightarrow Y$. Therefore, it admits a global section

$$\Gamma : Y \rightarrow \text{Ker } \widehat{L} \subset J^1(X, Y), \quad a_{ij}^{\alpha\mu} \Gamma_\mu^j + b_i^\alpha = 0, \quad (5.305)$$

which is a connection on $Y \rightarrow X$. If the Lagrangian (5.268) is regular, the connection (5.305) is unique.

There exists a linear bundle map

$$\sigma : \Pi \longrightarrow T^*X \otimes VY, \quad \bar{y}_\alpha^i \circ \sigma = \sigma_{\alpha\mu}^{ij} p_j^\mu, \quad (5.306)$$

such that

$$a \circ \sigma \circ a = a, \quad a_{ij}^{\alpha\mu} \sigma_{\mu\alpha}^{jk} a_{kb}^{\alpha\nu} = a_{ib}^{\alpha\nu}. \quad (5.307)$$

Note that σ is not unique, but it falls into the sum $\sigma = \sigma_0 + \sigma_1$ such that

$$\sigma_0 \circ a \circ \sigma_0 = \sigma_0, \quad a \circ \sigma_1 = \sigma_1 \circ a = 0, \quad (5.308)$$

where σ_0 is uniquely defined. For example, there exists a nondegenerate map σ (5.306).

Recall that there are the splittings

$$\begin{aligned} J^1(X, Y) &= \mathcal{S}(J^1(X, Y)) \oplus \mathcal{F}(J^1(X, Y)) = \text{Ker } \widehat{L} \oplus \text{Im}(\sigma_0 \circ \widehat{L}) \quad (5.309) \\ y_\alpha^i &= \mathcal{S}_\alpha^i + \mathcal{F}_\alpha^i = [y_\alpha^i - \sigma_{0\alpha\mu}^{ik} (a_{kj}^{\alpha\mu} y_\mu^j + b_k^\alpha)] + [\sigma_{0\alpha\mu}^{ik} (a_{kj}^{\alpha\mu} y_\mu^j + b_k^\alpha)], \\ \Pi &= \mathcal{R}(\Pi) \oplus \mathcal{P}(\Pi) = \text{Ker } \sigma_0 \oplus N_L, \quad (5.310) \\ p_i^\alpha &= \mathcal{R}_i^\alpha + \mathcal{P}_i^\alpha = [p_i^\alpha - a_{ij}^{\alpha\mu} \sigma_{0\mu\alpha}^{jk} p_k^\alpha] + [a_{ij}^{\alpha\mu} \sigma_{0\mu\alpha}^{jk} p_k^\alpha]. \end{aligned}$$

The relations (5.308) lead to the equalities

$$a_{ij}^{\alpha\mu} \mathcal{S}_\mu^j = 0, \quad \sigma_{0\mu\alpha}^{jk} \mathcal{R}_k^\alpha = 0, \quad \sigma_{1\mu\alpha}^{jk} \mathcal{P}_k^\alpha = 0, \quad \mathcal{R}_i^\alpha \mathcal{F}_\alpha^i = 0. \quad (5.311)$$

Due to these equalities, the Lagrangian (5.268) takes the form

$$L = \mathcal{L}\omega, \quad \mathcal{L} = \frac{1}{2} a_{ij}^{\alpha\mu} \mathcal{F}_\alpha^i \mathcal{F}_\mu^j + c'. \quad (5.312)$$

One can show that, this Lagrangian admits a set of associated Hamiltonians

$$\mathcal{H}_\Gamma = (\mathcal{R}_i^\alpha + \mathcal{P}_i^\alpha) \Gamma_\alpha^i + \frac{1}{2} \sigma_{0\alpha\mu}^{ij} \mathcal{P}_i^\alpha \mathcal{P}_j^\mu + \frac{1}{2} \sigma_{1\alpha\mu}^{ij} \mathcal{R}_i^\alpha \mathcal{R}_j^\mu - c' \quad (5.313)$$

indexed by connections Γ (5.305). Therefore, the Lagrangian constraint manifold (5.303) is given by the reducible constraints

$$\mathcal{R}_i^\alpha = p_i^\alpha - a_{ij}^{\alpha\mu} \sigma_{0\mu\alpha}^{jk} p_k^\alpha = 0. \quad (5.314)$$

Given a Hamiltonian \mathcal{H}_Γ , the corresponding Lagrangian (5.276) reads

$$\mathcal{L}_{\mathcal{H}_\Gamma} = \mathcal{R}_i^\alpha (\mathcal{S}_\alpha^i - \Gamma_\alpha^i) + \mathcal{P}_i^\alpha \mathcal{F}_\alpha^i - \frac{1}{2} \sigma_{0\alpha\mu}^{ij} \mathcal{P}_i^\alpha \mathcal{P}_j^\mu - \frac{1}{2} \sigma_{1\alpha\mu}^{ij} \mathcal{R}_i^\alpha \mathcal{R}_j^\mu + c'. \quad (5.315)$$

Its restriction (5.277) to the Lagrangian constraint manifold N_L (5.314) is

$$L_N = \mathcal{L}_N \omega, \quad \mathcal{L}_N = \mathcal{P}_i^\alpha \mathcal{F}_\alpha^i - \frac{1}{2} \sigma_{0\alpha\mu}^{ij} \mathcal{P}_i^\alpha \mathcal{P}_j^\mu + c'. \quad (5.316)$$

It is independent of the choice of a Hamiltonian (5.313). Note that the Lagrangian \mathcal{L}_N may admit gauge symmetries due to the term $\mathcal{P}_i^\alpha \mathcal{F}_\alpha^i$.

The Hamiltonian \mathcal{H}_Γ induces the Hamiltonian map \widehat{H}_Γ (5.280) and the projector

$$T = \widehat{L} \circ \widehat{H}_\Gamma, \quad p_i^\alpha \circ T = T_{i\mu}^{\alpha j} p_j^\mu = a_{ik}^{\alpha\nu} \sigma_{0\nu\mu}^{kj} p_j^\mu = \mathcal{P}_i^\alpha, \quad (5.317)$$

from Π onto its summand N_L in the decomposition (5.310). It obeys the relations

$$\sigma \circ T = \sigma_0, \quad T \circ a = a. \quad (5.318)$$

The projector T (5.317) is a linear map over Id_Y . Therefore, $T : \Pi \rightarrow N_L$ is a vector bundle. Let us consider the pull-back $L_\Pi = T^*L_N$ of the constrained Lagrangian L_N (5.316) onto Π . Due to the relations (5.311), it is given by the coordinate expression

$$L_\Pi = \mathcal{L}_\Pi \omega, \quad \mathcal{L}_\Pi = p_i^\alpha \mathcal{F}_\alpha^i - \frac{1}{2} \sigma_{0\alpha\mu}^{ij} p_i^\alpha p_j^\mu + c'. \quad (5.319)$$

This Lagrangian is gauge-invariant under the subgroup of the gauge group of vertical automorphisms Φ of the affine bundle $\Pi \rightarrow Y$ such that $T \circ \Phi = T$. This subgroup coincides with the gauge group $\text{Aut Ker } \sigma_0$ of vertical automorphisms of the vector bundle $\text{Ker } \sigma_0 \rightarrow Y$.

Note that the splittings (5.309) and (5.310) result from the splitting of the vector bundle

$$T^*X \otimes VY = \text{Ker } a \oplus E,$$

which can be provided with the adapted coordinates (\bar{y}^a, \bar{y}^A) such that a (5.304) is brought into a diagonal matrix with non-vanishing components a_{AA} . Then the Legendre bundle $\Pi \rightarrow Y$ (5.283) admits the dual (nonholonomic) coordinates (p_a, p_A) , where p_A are coordinates on the Lagrangian constraint manifold N_L , given by the irreducible constraints $p_a = 0$. Written relative to these coordinates, σ_0 becomes the diagonal matrix

$$\sigma_0^{AA} = (a_{AA})^{-1}, \quad \sigma_0^{aa} = 0, \quad (5.320)$$

while $\sigma_1^{Aa} = \sigma_1^{AB} = 0$. Moreover, one can choose the coordinates \bar{y}^a (accordingly, p_a) and the map σ (5.306) such that σ_1 becomes a diagonal matrix with non-vanishing positive components $\sigma_1^{aa} = \mathcal{V}^{-1}$, where $\mathcal{V}\omega$ is a volume form on X . We further follow this choice of the adapted coordinates (p_a, p_A) . Let us write

$$p_a = M_{a\alpha}^i p_i^\alpha, \quad p_A = M_{A\alpha}^i p_i^\alpha, \quad (5.321)$$

where M are the matrix functions on Y obeying the relations

$$M_{a\alpha}^i a_{ij}^{\alpha\mu} = 0, \quad M^{-1\alpha a}_i \sigma_{0\alpha}^i = 0, \quad M_{a\alpha}^i \mathcal{P}_i^\alpha = 0, \quad M_{A\alpha}^i \mathcal{R}_i^\alpha = 0.$$

Then the Lagrangian L_N (5.316) with respect to the adapted coordinates (p_a, p_A) takes the form

$$\mathcal{L}_N = M^{-1\alpha A}_i p_A \mathcal{F}_\alpha^i - \frac{1}{2} \sum_A (a_{AA})^{-1} (p_A)^2 + c', \quad (5.322)$$

5.11 Application: Gauge Fields on Principal Connections

5.11.1 Connection Strength

Given a principal G -bundle $P \rightarrow Q$, the *Frölicher–Nijenhuis bracket* (4.140) on the space $\wedge^*(P) \otimes \mathcal{V}^1(P)$ of tangent-valued forms on P is compatible with the canonical action R_G (4.31) of G on P , and induces the Frölicher–Nijenhuis bracket on the space $\wedge^*(Q) \otimes T_GP(Q)$ of T_GP -valued forms on Q . Its coordinate form issues from the Lie bracket (4.36).

Then any principal connection $A \in \wedge^1(Q) \otimes T_GP(Q)$ (5.42) sets the *Nijenhuis differential*

$$\begin{aligned} d_A : \wedge^r(Q) \otimes T_GP(Q) &\rightarrow \wedge^{r+1}(Q) \otimes V_GP(Q), \\ d_A\phi &= [A, \phi]_{FN}, \quad \phi \in \wedge^r(Q) \otimes T_GP(Q), \end{aligned} \quad (5.323)$$

on the space $\wedge^*(Q) \otimes T_GP(Q)$.

The curvature R (5.34) can be equivalently defined as the Nijenhuis differential

$$R : Y \rightarrow \wedge^2 T^*Q \otimes VY, \quad \text{given by} \quad R = \frac{1}{2} d_\Gamma \Gamma = \frac{1}{2} [\Gamma, \Gamma]_{FN}. \quad (5.324)$$

Let us define the *strength* of a principal connection A , as

$$F_A = \frac{1}{2} d_A A = \frac{1}{2} [A, A]_{FN} \in \wedge^2(Q) \otimes V_GP(Q), \quad (5.325)$$

$$F_A = \frac{1}{2} F_{\lambda\mu}^r dq^\alpha \wedge dq^\mu \otimes e_r, \quad F_{\lambda\mu}^r = \partial_\alpha A_\mu^r - \partial_\mu A_\alpha^r + c_{pq}^r A_\alpha^p A_\mu^q. \quad (5.326)$$

It is locally given by the expression

$$F_A = d\mathbf{A} + \frac{1}{2} [\mathbf{A}, \mathbf{A}] = d\mathbf{A} + \mathbf{A} \wedge \mathbf{A}, \quad (5.327)$$

where \mathbf{A} is the local connection form (5.43). By definition, the strength (5.325) of a principal connection obeys the second *Bianchi identity*

$$d_A F_A = [A, F_A]_{FN} = 0. \quad (5.328)$$

It should be emphasized that the strength F_A (5.325) is not the standard curvature (5.34) of a connection on P , but there are the local relations $\psi_\zeta^P F_A = z_\zeta^* \Theta$, where

$$\Theta = d\tilde{A} + \frac{1}{2} [\tilde{A}, \tilde{A}] \quad (5.329)$$

is the \mathfrak{g}_l -valued curvature form on P (see the expression (5.333) below). In particular, a principal connection is flat iff its strength vanishes.

5.11.2 Associated Bundles

Given a principal G -bundle $\pi_P : P \rightarrow Q$, let V be a manifold provided with an effective left action $G \times V \rightarrow V$, $(g, v) \mapsto gv$ of the Lie group G . Let us consider the quotient

$$Y = (P \times V)/G \quad (5.330)$$

of the product $P \times V$ by identification of elements (p, v) and $(pg, g^{-1}v)$ for all $g \in G$. We will use the notation $(pG, G^{-1}v)$ for its points. Let $[p]$ denote the restriction of the canonical surjection

$$P \times V \rightarrow (P \times V)/G \quad (5.331)$$

to the subset $\{p\} \times V$ so that $[p](v) = [pg](g^{-1}v)$. Then the map $Y \rightarrow Q$, $[p](V) \mapsto \pi_P(p)$, makes the quotient Y (5.330) to a fibre bundle over Q .

Let us note that, for any G -bundle, there exists an associated principal G -bundle [Steenrod (1951)]. The peculiarity of the G -bundle Y (5.330) is that it appears canonically associated to a principal bundle P . Indeed, every bundle atlas $\Psi_P = \{(U_\alpha, z_\alpha)\}$ of P determines a unique associated bundle atlas

$$\Psi = \{(U_\alpha, \psi_\alpha(q) = [z_\alpha(q)]^{-1})\}$$

of the quotient Y (5.330), and each automorphism of P also induces the corresponding automorphism (5.346) of Y .

Every principal connection A on a principal bundle $P \rightarrow Q$ induces a unique connection on the associated fibre bundle Y (5.330). Given the horizontal splitting of the tangent bundle TP relative to A , the tangent map to the canonical map (5.331) defines the horizontal splitting of the tangent bundle TY of Y and, consequently, a connection on $Y \rightarrow Q$ [Kobayashi and Nomizu (1963/9)]. This is called the associated principal connection or a principal connection on a P -associated bundle $Y \rightarrow Q$. If Y is a vector bundle, this connection takes the form

$$A = dq^\alpha \otimes (\partial_\alpha - A_\alpha^p I_{pj}^i y^j \partial_i), \quad (5.332)$$

where I_p are generators of the linear representation of the Lie algebra \mathfrak{g}_r in

the vector space V . The curvature (5.34) of this connection reads

$$R = -\frac{1}{2}F_{\lambda\mu}^p I_{pj}^i y^j dq^\alpha \wedge dq^\mu \otimes \partial_i, \quad (5.333)$$

where $F_{\lambda\mu}^p$ are coefficients (5.326) of the strength of a principal connection A .

In particular, any principal connection A induces the associated linear connection on the gauge algebra bundle $V_G P \rightarrow Q$. The corresponding covariant differential $\nabla^A \xi$ (5.30) of its sections $\xi = \xi^p e_p$ reads

$$\nabla^A \xi : Q \rightarrow T^*Q \otimes V_G P, \quad \nabla^A \xi = (\partial_\alpha \xi^r + c_{pq}^r A_\alpha^p \xi^q) dq^\alpha \otimes e_r.$$

It coincides with the Nijenhuis differential

$$d_A \xi = [A, \xi]_{FN} = \nabla^A \xi \quad (5.334)$$

of ξ seen as a $V_G P$ -valued 0-form, and is given by the local expression given by the local expression

$$\nabla^A \xi = d\xi + [\mathbf{A}, \xi], \quad (5.335)$$

where \mathbf{A} is the local connection form (5.43).

5.11.3 Classical Gauge Fields

Since gauge potentials are represented by global sections of the connection bundle $C \rightarrow Q$ (5.45), its 1-jet space $J^1(Q, C)$ plays the role of a configuration space of classical gauge theory. The key point is that the jet space $J^1(Q, C)$ admits the canonical splitting over C which leads to a unique canonical *Yang–Mills Lagrangian density* of gauge theory on $J^1(Q, C)$.

Let us describe this splitting. One can show that the principal G -bundle

$$J^1(Q, P) \rightarrow J^1(Q, P)/G = C \quad (5.336)$$

is canonically isomorphic to the trivial pull-back bundle

$$P_C = C \times P \rightarrow C, \quad (5.337)$$

and that the latter admits the canonical principal connection [García (1977); Giachetta *et. al.* (1997)]

$$\mathcal{A} = dq^\alpha \otimes (\partial_\alpha + a_\alpha^p e_p) + da_\alpha^r \otimes \partial_r^\alpha \in \mathcal{O}^1(C) \otimes T_G(P_C)(C). \quad (5.338)$$

Since C (5.45) is an affine bundle modelled over the vector bundle

$$\overline{C} = T^*Q \otimes V_G P \rightarrow Q,$$

the vertical tangent bundle of C possesses the canonical trivialization

$$\begin{aligned} VC &= C \times T^*Q \otimes V_G P, & \text{where} \\ V_G P_C &= V_G(C \times P) = C \times V_G P. \end{aligned} \quad (5.339)$$

Then the strength F_A of the connection (5.338) is the $V_G P$ -valued horizontal 2-form on C ,

$$\begin{aligned} F_A &= \frac{1}{2} d_A \mathcal{A} = \frac{1}{2} [\mathcal{A}, \mathcal{A}]_{FN} \in \wedge^2(C) \otimes V_G P(Q), \\ F_A &= (da_\mu^r \wedge dq^\mu + \frac{1}{2} c_{pq}^r a_\alpha^p a_\mu^q dq^\alpha \wedge dq^\mu) \otimes e_r. \end{aligned} \quad (5.340)$$

Note that, given a global section connection A of the connection bundle $C \rightarrow Q$, the pull-back $A^* F_A = F_A$ is the strength (5.325) of the principal connection A .

Let us take the pull-back of the form F_A onto $J^1(Q, C)$ with respect to the fibration (5.336), and consider the $V_G P$ -valued horizontal 2-form

$$\begin{aligned} \mathcal{F} &= h_0(F_A) = \frac{1}{2} \mathcal{F}_{\lambda\mu}^r dq^\alpha \wedge dq^\mu \otimes e_r, \\ \mathcal{F}_{\lambda\mu}^r &= a_{\lambda\mu}^r - a_{\mu\lambda}^r + c_{pq}^r a_\alpha^p a_\mu^q, \end{aligned} \quad (5.341)$$

where h_0 is the horizontal projection (5.58). Note that

$$\mathcal{F}/2 : J^1(Q, C) \rightarrow C \times \wedge^2 T^*Q \otimes V_G P \quad (5.342)$$

is an affine map over C of constant rank. Hence, its kernel $C_+ = \text{Ker } \mathcal{F}$ is the affine subbundle of $J^1(Q, C) \rightarrow C$, and we have a desired canonical splitting

$$J^1(Q, C) = C_+ \oplus C_- = C_+ \oplus (C \times \wedge^2 T^*Q \otimes V_G P), \quad (5.343)$$

$$a_{\lambda\mu}^r = \frac{1}{2} (a_{\lambda\mu}^r + a_{\mu\lambda}^r - c_{pq}^r a_\alpha^p a_\mu^q) + \frac{1}{2} (a_{\lambda\mu}^r - a_{\mu\lambda}^r + c_{pq}^r a_\alpha^p a_\mu^q), \quad (5.344)$$

over C of the jet space $J^1(Q, C)$. The corresponding canonical projections are

$$\pi_1 = \mathcal{S} : J^1(Q, C) \rightarrow C_+, \quad \mathcal{S}_{\lambda\mu}^r = \frac{1}{2} (a_{\lambda\mu}^r + a_{\mu\lambda}^r - c_{pq}^r a_\alpha^p a_\mu^q), \quad (5.345)$$

and $\pi_2 = \mathcal{F}/2$ given by (5.342).

5.11.4 Gauge Transformations

In classical gauge theory, several classes of gauge transformations are examined [Giachetta *et. al.* (1997); Marathe and Martucci (1992); Socolovsky (1991)]. A most general gauge transformation is defined as an automorphism Φ_P of a principal G -bundle P which is equivariant under the canonical action (4.31) of the group G on G , i.e.,

$$R_g \circ \Phi_P = \Phi_P \circ R_g, \quad (g \in G).$$

Such an automorphism of P induces the corresponding automorphism

$$\Phi_Y : (pG, G^{-1}v) \rightarrow (\Phi_P(p)G, G^{-1}v) \quad (5.346)$$

of the P -associated bundle Y (5.330) and the corresponding automorphism

$$\Phi_C : J^1(Q, P)/G \rightarrow J^1\Phi_P(J^1(Q, P))/G \quad (5.347)$$

of the connection bundle C .

Every vertical automorphism of a principal bundle P is represented as

$$\Phi_P(p) = pf(p), \quad (p \in P), \quad (5.348)$$

where f is a G -valued equivariant function on P , i.e.,

$$f(pg) = g^{-1}f(p)g, \quad (g \in G). \quad (5.349)$$

There is a 1-1 correspondence between these functions and the global sections s of the group bundle

$$P^G = (P \times G)/G, \quad (5.350)$$

whose typical fibre is the group G , subject to the adjoint representation of the structure group G . Therefore, P^G (5.350) is also called the *adjoint bundle*. There is the canonical fibre-to-fibre action of the group bundle P^G on any P -associated bundle Y by the law

$$P^G \times Y \rightarrow Y, \quad ((pG, G^{-1}gG), (pG, G^{-1}v)) \mapsto (pG, G^{-1}gv).$$

Then the above-mentioned correspondence is given by the relation

$$P^G \times P \rightarrow P, \quad (s(\pi_P(p)), p) \mapsto pf(p),$$

where P is a G -bundle associated to itself. Hence, the gauge group $\mathfrak{G}(P)$ of vertical automorphisms of a principal G -bundle $P \rightarrow Q$ is isomorphic to the group of global sections of the P -associated group bundle (5.350).

In order to study the gauge invariance of one or another object in gauge theory, it suffices to examine its invariance under an arbitrary 1-parameter subgroup $[\Phi_P]$ of the gauge group. Its infinitesimal generator is a G -invariant vertical vector-field ξ on a principal bundle P or, equivalently, a section

$$\xi = \xi^p(x)e_p \quad (5.351)$$

of the gauge algebra bundle $V_GP \rightarrow Q$ (4.35). We will call it a *gauge vector-field*. One can think of its components $\xi^p(q)$ as being gauge parameters. Gauge vector-fields (5.351) are transformed under the infinitesimal generators of gauge transformations (i.e., other gauge vector-fields) ξ' by the adjoint representation

$$\mathfrak{L}_{\xi'}\xi = [\xi', \xi] = c_{rq}^p \xi'^r \xi^q e_p, \quad (\xi, \xi' \in V_GP(Q)).$$

Therefore, gauge parameters are subject to the coadjoint representation

$$\xi' : \xi^p \mapsto -c_{rq}^p \xi'^r \xi^q. \quad (5.352)$$

Given a gauge vector-field ξ (5.351) seen as the infinitesimal generator of a 1-parameter gauge group $[\Phi_P]$, let us get the gauge vector-fields on a P -associated bundle Y and the connection bundle C .

The corresponding gauge vector-field on the P -associated vector bundle $Y \rightarrow Q$ issues from the relation (5.346), and reads

$$\xi_Y = \xi^p I_p^i \partial_i,$$

where I_p are generators of the group G , acting on the typical fibre V of Y .

The gauge vector-field ξ (5.351) acts on elements a (5.46) of the connection bundle by the law

$$\mathfrak{L}_\xi a = [\xi, a]_{FN} = (-\partial_\alpha \xi^r + c_{pq}^r \xi^p a_\alpha^q) dq^\alpha \otimes e_r.$$

In view of the vertical splitting (5.339), this quantity can be regarded as the vertical vector-field

$$\xi_C = (-\partial_\alpha \xi^r + c_{pq}^r \xi^p a_\alpha^q) \partial_r \quad (5.353)$$

on the connection bundle C , and is the infinitesimal generator of the 1-parameter group $[\Phi_C]$ of vertical automorphisms (5.347) of C , i.e., a desired gauge vector-field on C .

5.11.5 Lagrangian Gauge Theory

Classical gauge theory of unbroken symmetries on a principal G -bundle $P \rightarrow Q$ deals with two types of fields. These are gauge potentials identified to global sections of the connection bundle $C \rightarrow Q$ (5.45) and matter fields represented by global sections of a P -associated vector bundle Y (5.330), called a matter bundle. Therefore, the total configuration space of classical gauge theory is the product of jet bundles

$$J^1(X, Y)_{tot} = J^1(X, Y) \times J^1(Q, C). \quad (5.354)$$

Let us study a gauge-invariant Lagrangian on this configuration space.

A total gauge vector-field on the product $C \times Y$ reads

$$\xi_{YC} = (-\partial_\alpha \xi^r + c_{pq}^r \xi^p a_\alpha^q) \partial_r^\alpha + \xi^p I_p^i \partial_i = (u_p^{A\lambda} \partial_\alpha \xi^p + u_p^A \xi^p) \partial_A, \quad (5.355)$$

where we use the collective index A , and put the notation

$$u_p^{A\lambda} \partial_A = -\delta_p^r \partial_r^\alpha, \quad u_p^A \partial_A = c_{qp}^r a_\alpha^q \partial_r^\alpha + I_p^i \partial_i.$$

A Lagrangian L on the configuration space (5.354) is said to be gauge-invariant if its Lie derivative $\mathfrak{L}_{J^1 \xi_{YC}} L$ along any gauge vector-field ξ (5.351) vanishes. Then the *first variational formula* (5.109) leads to the strong equality

$$0 = (u_p^A \xi^p + u_p^{A\mu} \partial_\mu \xi^p) \delta_A \mathcal{L} + d_\alpha [(u_p^A \xi^p + u_p^{A\mu} \partial_\mu \xi^p) \pi_A^\alpha], \quad (5.356)$$

where $\delta_A \mathcal{L}$ are the variational derivatives (5.233) of L and the total derivative reads

$$d_\alpha = \partial_\alpha + a_{\lambda\mu}^p \partial_p^\mu + y_\alpha^i \partial_i.$$

Due to the arbitrariness of gauge parameters ξ^p , this equality falls into the system of strong equalities

$$u_p^A \delta_A \mathcal{L} + d_\mu (u_p^A \pi_A^\mu) = 0, \quad (5.357)$$

$$u_p^{A\mu} \delta_A \mathcal{L} + d_\alpha (u_p^{A\mu} \pi_A^\alpha) + u_p^A \pi_A^\mu = 0, \quad (5.358)$$

$$u_p^{A\lambda} \pi_A^\mu + u_p^{A\mu} \pi_A^\lambda = 0. \quad (5.359)$$

Substituting (5.358) and (5.359) in (5.357), we get the well-known constraints

$$u_p^A \delta_A \mathcal{L} - d_\mu (u_p^{A\mu} \delta_A \mathcal{L}) = 0$$

for the variational derivatives of a gauge-invariant Lagrangian L .

Treating the equalities (5.357) – (5.359) as the equations for a gauge-invariant Lagrangian, let us solve these equations for a Lagrangian

$$L = \mathcal{L}(t, q^i, a_{\mu}^r, a_{\lambda\mu}^r) \omega : J^1(Q, C) \rightarrow \wedge^n T^*Q \quad (5.360)$$

without matter fields. In this case, the equations (5.357) – (5.359) read

$$c_{pq}^r (a_{\mu}^p \partial_r^{\mu} \mathcal{L} + a_{\lambda\mu}^p \partial_r^{\lambda\mu} \mathcal{L}) = 0, \quad (5.361)$$

$$\partial_q^{\mu} \mathcal{L} + c_{pq}^r a_{\alpha}^p \partial_r^{\mu\lambda} \mathcal{L} = 0, \quad (5.362)$$

$$\partial_p^{\mu\lambda} \mathcal{L} + \partial_p^{\lambda\mu} \mathcal{L} = 0. \quad (5.363)$$

Let rewrite them relative to the coordinates $(a_{\mu}^q, S_{\mu\lambda}^r, \mathcal{F}_{\mu\lambda}^r)$ (5.341) and (5.345), associated to the canonical splitting (5.343) of the jet space $J^1(Q, C)$. The equation (5.363) reads

$$\frac{\partial \mathcal{L}}{\partial S_{\mu\lambda}^r} = 0. \quad (5.364)$$

Then a simple computation brings the equation (5.362) into the form

$$\partial_q^{\mu} \mathcal{L} = 0. \quad (5.365)$$

The equations (5.364) and (5.365) shows that the gauge-invariant Lagrangian (5.360) factorizes through the strength \mathcal{F} (5.341) of gauge potentials. As a consequence, the equation (5.361) takes the form

$$c_{pq}^r \mathcal{F}_{\lambda\mu}^p \frac{\partial \mathcal{L}}{\partial \mathcal{F}_{\lambda\mu}^r} = 0.$$

It admits a unique solution in the class of quadratic Lagrangians which is the conventional *Yang–Mills Lagrangian* L_{YM} of gauge potentials on the configuration space $J^1(Q, C)$. In the presence of a background world metric g on the base Q , it reads

$$L_{YM} = \frac{1}{4\varepsilon^2} a_{pq}^G g^{\lambda\mu} g^{\beta\nu} \mathcal{F}_{\lambda\beta}^p \mathcal{F}_{\mu\nu}^q \sqrt{|g|} \omega, \quad (\text{where } g = \det(g_{\mu\nu})), \quad (5.366)$$

where a^G is a G -invariant bilinear form on the Lie algebra of \mathfrak{g}_r and ε is a coupling constant.

5.11.6 Hamiltonian Gauge Theory

Let us consider gauge theory of principal connections on a principal bundle $P \rightarrow X$ with a structure Lie group G . Principal connections on $P \rightarrow X$

are represented by sections of the affine bundle

$$C = J^1(Q, P)/G \rightarrow X, \quad (5.367)$$

modelled over the vector bundle $T^*X \otimes V_GP$ [Giachetta *et. al.* (1997)]. Here, $V_GP = VP/G$ is the fibre bundle in Lie algebras \mathfrak{g} of the group G . Given the basis $\{\varepsilon_r\}$ for \mathfrak{g} , we get the local fibre bases $\{e_r\}$ for V_GP . The connection bundle C (5.367) is coordinated by (x^μ, a_μ^r) such that, written relative to these coordinates, sections $A = A_\mu^r dx^\mu \otimes e_r$ of $C \rightarrow X$ are the familiar local connection 1-forms, regarded as *gauge potentials*.

There is 1–1 correspondence between the sections $\xi = \xi^r e_r$ of $V_GP \rightarrow X$ and the vector-fields on P which are infinitesimal generators of 1-parameter groups of vertical automorphisms (gauge transformations) of P . Any section ξ of $V_GP \rightarrow X$ induces the vector-field on C , given by

$$u(\xi) = u_\mu^k \frac{\partial}{\partial a_\mu^r} = (c_{pq}^r a_\mu^p \xi^q + \partial_\mu \xi^r) \frac{\partial}{\partial a_\mu^r}, \quad (5.368)$$

where c_{pq}^r are the structure constants of the Lie algebra \mathfrak{g} .

The configuration space of gauge theory is the jet space $J^1(Q, C)$ equipped with the coordinates $(x^\alpha, a_\alpha^m, a_{\mu\lambda}^m)$. It admits the canonical splitting (5.309) given by the coordinate expression

$$a_{\mu\lambda}^r = \mathcal{S}_{\mu\lambda}^r + \mathcal{F}_{\mu\lambda}^r = \frac{1}{2}(a_{\mu\lambda}^r + a_{\lambda\mu}^r - c_{pq}^r a_\mu^p a_\lambda^q) + \frac{1}{2}(a_{\mu\lambda}^r - a_{\lambda\mu}^r + c_{pq}^r a_\mu^p a_\lambda^q),$$

where \mathcal{F} is the strength of gauge fields up to the factor 1/2. The *Yang–Mills Lagrangian* L_{YM} on the configuration space $J^1(Q, C)$ is given by

$$L_{YM} = a_{pq}^G g^{\lambda\mu} g^{\beta\nu} \mathcal{F}_{\lambda\beta}^p \mathcal{F}_{\mu\nu}^q \sqrt{|g|} \omega, \quad (g = \det(g_{\mu\nu})), \quad (5.369)$$

where a^G is a non-degenerate G -invariant metric in the dual of the Lie algebra of \mathfrak{g} and g is a pseudo-Riemannian metric on X .

The phase-space Π (5.271) of the gauge theory is with the canonical coordinates $(x^\alpha, a_\alpha^p, p_q^{\mu\lambda})$. It admits the canonical splitting (5.310) given by the coordinate expression

$$p_m^{\mu\lambda} = \mathcal{R}_m^{\mu\lambda} + \mathcal{P}_m^{\mu\lambda} = p_m^{(\mu\lambda)} + p_m^{[\mu\lambda]} = \frac{1}{2}(p_m^{\mu\lambda} + p_m^{\lambda\mu}) + \frac{1}{2}(p_m^{\mu\lambda} - p_m^{\lambda\mu}). \quad (5.370)$$

With respect to this splitting, the Legendre map induced by the Lagrangian (5.369) takes the form

$$p_m^{(\mu\lambda)} \circ \widehat{L}_{YM} = 0, \quad (5.371)$$

$$p_m^{[\mu\lambda]} \circ \widehat{L}_{YM} = 4a_{mn}^G g^{\mu\alpha} g^{\lambda\beta} \mathcal{F}_{\alpha\beta}^n \sqrt{|g|}. \quad (5.372)$$

The equalities (5.371) define the Lagrangian constraint space N_L of Hamiltonian gauge theory. Obviously, it is an imbedded submanifold of Π , and the Lagrangian L_{YM} is almost-regular.

In order to construct an associated Hamiltonian, let us consider a connection Γ (5.305) on the fibre bundle $C \rightarrow X$ which take their values into $\text{Ker } \widehat{L}$, i.e.,

$$\Gamma_{\lambda\mu}^r - \Gamma_{\mu\lambda}^r + c_{pq}^r a_\lambda^p a_\mu^q = 0.$$

Given a symmetric linear connection K on X and a principal connection B on $P \rightarrow X$, this connection reads

$$\Gamma_{\lambda\mu}^r = \frac{1}{2}[\partial_\mu B_\alpha^r + \partial_\alpha B_\mu^r - c_{pq}^r a_\alpha^p a_\mu^q + c_{pq}^r (a_\alpha^p B_\mu^q + a_\mu^p B_\alpha^q)] - K_\alpha{}^\beta{}_\mu (a_\beta^r - B_\beta^r).$$

The corresponding Hamiltonian (5.313) associated to L_{YM} is

$$\mathcal{H}_\Gamma = p_r^{\lambda\mu} \Gamma_{\lambda\mu}^r + a_G^{mn} g_{\mu\nu} g_{\lambda\beta} p_m^{[\mu\lambda]} p_n^{[\nu\beta]} \sqrt{|g|}.$$

Then we get the Lagrangian

$$\mathcal{L}_N = p_r^{[\lambda\mu]} \mathcal{F}_{\lambda\mu}^r - a_G^{mn} g_{\mu\nu} g_{\lambda\beta} p_m^{[\mu\lambda]} p_n^{[\nu\beta]} \sqrt{|g|}$$

(5.316) on the Lagrangian constraint manifold (5.371) and its pull-back

$$L_\Pi = \mathcal{L}_\Pi \omega, \quad \mathcal{L}_\Pi = p_r^{\lambda\mu} \mathcal{F}_{\lambda\mu}^r - a_G^{mn} g_{\mu\nu} g_{\lambda\beta} p_m^{[\mu\lambda]} p_n^{[\nu\beta]} \sqrt{|g|}, \quad (5.373)$$

(5.319) onto Π .

Both the Lagrangian L_{YM} (5.369) on C and the Lagrangian L_Π (5.373) on Π are invariant under gauge transformations whose infinitesimal generators are the lifts

$$\begin{aligned} j^1 u(\xi) &= (c_{pq}^r a_\mu^p \xi^q + \partial_\mu \xi^r) \frac{\partial}{\partial a_\mu^r} + (c_{pq}^r (a_{\lambda\mu}^p \xi^q + a_\mu^p \partial_\alpha \xi^q) + \partial_\alpha \partial_\mu \xi^r) \frac{\partial}{\partial a_{\lambda\mu}^r}, \\ \bar{u}(\xi) &= j^1 u(\xi) - c_{pq}^r p_r^{\lambda\mu} \xi^q \frac{\partial}{\partial p_p^{\lambda\mu}}, \end{aligned}$$

of the vector-fields (5.368) onto $J^1(Q, C)$ and $\Pi \times J^1(Q, C)$, respectively. We have the transformation laws

$$j^1 u(\xi)(\mathcal{F}_{\lambda\mu}^r) = c_{pq}^r \mathcal{F}_{\lambda\mu}^p \xi^q, \quad j^1 u(\xi)(\mathcal{S}_{\lambda\mu}^r) = c_{pq}^r \mathcal{S}_{\lambda\mu}^p \xi^q + c_{pq}^r a_\mu^p \partial_\alpha \xi^q + \partial_\alpha \partial_\mu \xi^r.$$

Therefore, one can choose the gauge conditions

$$g^{\lambda\mu} \mathcal{S}_{\lambda\mu}^r(x) - \alpha^r(x) = \frac{1}{2} g^{\lambda\mu} (\partial_\alpha a_\mu^r(x) + \partial_\mu a_\alpha^r(x)) - \alpha^r(x) = 0,$$

which are the familiar generalized Lorentz gauge. The corresponding second-order differential operator (6.95) reads

$$M_s^r \xi^s = g^{\lambda\mu} \left(\frac{1}{2} c_{pq}^r (\partial_\alpha a_\mu^r + \partial_\mu a_\alpha^r) \xi^q + c_{pq}^r a_\mu^p \partial_\alpha \xi^q + \partial_\alpha \partial_\mu \xi^r \right).$$

Passing to the Euclidean space and repeating the above quantization procedure, we come to the generating functional

$$\begin{aligned} Z = \mathcal{N}^{-1} \int \exp \{ & \int (p_r^{\alpha\mu} \mathcal{F}_{\alpha\mu}^r - a_G^{mn} g_{\mu\nu} g_{\alpha\beta} p_m^{\mu\alpha} p_n^{\nu\beta} \sqrt{|g|} \\ & - \frac{1}{8} a_{rs}^G g^{\alpha\nu} g^{\alpha\mu} (\partial_\alpha a_\nu^r + \partial_\nu a_\alpha^r) (\partial_\alpha a_\mu^s + \partial_\mu a_\alpha^s) \\ & - g^{\alpha\mu} \bar{c}_r \left(\frac{1}{2} c_{pq}^r (\partial_\alpha a_\mu^r + \partial_\mu a_\alpha^r) c^q + c_{pq}^r a_\mu^p c_\alpha^q + c_{\alpha\mu}^r \right) \\ & + i J_r^\mu a_\mu^r + i J_{\mu\alpha}^r p_r^{\mu\alpha} \} \omega \} \prod_x [d\bar{c}][dc][dp(x)][da(x)]. \end{aligned}$$

Its integration with respect to momenta restarts the familiar generating functional of gauge theory.

5.11.7 Gauge Conservation Laws

On-shell, the strong equality (5.356) becomes the *weak Noether conservation law*

$$d_\alpha [(u_p^A \xi^p + u_p^{A\mu} \partial_\mu \xi^p) \pi_A^\alpha] \approx 0 \quad (5.374)$$

of the *Noether current*

$$\mathfrak{J}^\alpha = -(u_p^A \xi^p + u_p^{A\mu} \partial_\mu \xi^p) \pi_A^\alpha. \quad (5.375)$$

Therefore, the equalities (5.357) – (5.359) on-shell lead to the familiar *Noether identities*

$$d_\mu (u_p^A \pi_A^\mu) \approx 0, \quad (5.376)$$

$$d_\alpha (u_p^{A\mu} \pi_A^\mu) + u_p^A \pi_A^\mu \approx 0, \quad (5.377)$$

$$u_p^{A\alpha} \pi_A^\mu + u_p^{A\mu} \pi_A^\alpha = 0 \quad (5.378)$$

for a gauge-invariant Lagrangian L . They are equivalent to the weak equality (5.374) due to the arbitrariness of the gauge parameters $\xi^p(q)$.

The expressions (5.374) and (5.375) shows that both the Noether conservation law and the Noether current depend on gauge parameters. The weak identities (5.376) – (5.378) play the role of the necessary and sufficient

conditions in order that the Noether conservation law (5.374) is maintained under changes of gauge parameters. This means that, if the equality (5.374) holds for gauge parameters ξ , it does so for arbitrary deviations $\xi + \delta\xi$ of ξ . In particular, the Noether conservation law (5.374) is maintained under gauge transformations, when gauge parameters are transformed by the coadjoint representation (5.352).

It can be seen that the equalities (5.376) – (5.378) are not mutually independent, but (5.376) is a corollary of (5.377) and (5.378). This property reflects the fact that, in accordance with the strong equalities (5.358) and (5.359), the Noether current (5.375) is brought into the superpotential form

$$\mathfrak{J}^\alpha = \xi^p u_p^{A\alpha} \delta_A \mathcal{L} - d_\mu (\xi^p u_p^{A\mu} \pi_A^\alpha), \quad U^{\mu\alpha} = -\xi^p u_p^{A\mu} \pi_A^\alpha,$$

(5.232). Since a matter field Lagrangian is independent of the jet coordinates $a_{\alpha\mu}^p$, the *Noether superpotential*, $U^{\mu\alpha} = \xi^p \pi_p^{\mu\alpha}$, depends only on gauge potentials. The corresponding integral relation (5.234) reads

$$\int_{N^{n-1}} s^* \mathfrak{J}^\alpha \omega_\alpha = \int_{\partial N^{n-1}} s^* (\xi^p \pi_p^{\mu\alpha}) \omega_{\mu\alpha}, \quad (5.379)$$

where N^{n-1} is a compact oriented $(n-1)$ D submanifold of Q with the boundary ∂N^{n-1} . One can think of (5.379) as being the integral relation between the Noether current (5.375) and the gauge field, generated by this current. In electromagnetic theory seen as a $U(1)$ gauge theory, the similar relation between an electric current and the electromagnetic field generated by this current is well known. However, it is free from gauge parameters due to the peculiarity of Abelian gauge models.

Note that the Noether current (5.375) in gauge theory takes the superpotential form (5.232) because the infinitesimal generators of gauge transformations (5.355) depend on derivatives of gauge parameters.

5.11.8 Topological Gauge Theories

The field models that we have investigated so far show that when a background world metric is present, the stress–energy–momentum (SEM) transformation law becomes the covariant conservation law of the *metric SEM-tensor* (see subsection on SEM-tensors 5.12.1 below). Topological gauge theories exemplify the field models in the absence of a world metric [Giacchetta *et al.* (2005)].

Let us consider the *Chern–Simons gauge theory* on a 3D base manifold X^3 [Birmingham *et al.* (1991); Witten (1988a)]. Physical interpretation

of this fundamental model will be given in subsection 6.7.4 below.

Let $P \rightarrow X^3$ be a principal bundle with a structure semisimple Lie group G and C the corresponding bundle of principal connections which is coordinated by (x^α, k_α^m) . The *Chern–Simons Lagrangian density* is given by the coordinate expression

$$L_{CS} = \frac{1}{2k} a_{mn}^G \varepsilon^{\alpha\lambda\mu} k_\alpha^m (\mathcal{F}_{\lambda\mu}^n + \frac{1}{3} c_{pq}^n k_\alpha^p k_\mu^q) d^3x \quad (5.380)$$

where $\varepsilon^{\alpha\lambda\mu}$ is the skew-symmetric Levi–Civita tensor.

Note that the Lagrangian density (5.380) is not gauge-invariant and globally defined. At the same time, it gives the globally defined Euler–Lagrangian operator

$$\mathcal{E}_{L_{CS}} = \frac{1}{k} a_{mn}^G \varepsilon^{\alpha\lambda\mu} \mathcal{F}_{\lambda\mu}^n dk_\alpha^m \omega.$$

Thus, the gauge transformations in the Chern–Simons model appear to be the generalized invariant transformations which keep invariant the Euler–Lagrangian equations, but not the Lagrangian density. Solutions of these equations are the curvature-free principal connections A on the principal bundle $P \rightarrow X^3$.

Though the Chern–Simons Lagrangian density is not invariant under gauge transformations, we still have the Noether-type conservation law (5.374) in which the total conserved current is the standard Noether current (5.375) plus the additional term as follows.

Let $u_{\mathfrak{g}}$ be the principal vector-field (5.353) on the bundle of principal connections C . We calculate the Lie derivative

$$\mathfrak{L}_{u_{\mathfrak{g}}} L_{CS} = \frac{1}{k} a_{mn}^G \varepsilon^{\alpha\lambda\mu} \partial_\alpha (\alpha^m \partial_\alpha A_\mu^n) d^3x.$$

Hence, the Noether transformation law (5.374) becomes the conservation law

$$\begin{aligned} d_\alpha T_{CS}^\alpha &= d_\alpha (T^\alpha + \frac{1}{k} a_{mn}^G \varepsilon^{\alpha\lambda\mu} \alpha^m \partial_\alpha A_\mu^n) \approx 0, \quad \text{where} \quad (5.381) \\ T^\alpha &= \pi_n^{\mu\lambda} u_{\mathfrak{g}\mu}^n = \frac{1}{k} a_{mn}^G \varepsilon^{\alpha\lambda\mu} A_\alpha^m u_{\mathfrak{g}\mu}^n \end{aligned}$$

is the standard *Noether current*. After simplification, the conservation law (5.381) takes the form

$$d_\alpha (\frac{1}{k} a_{mn}^G \varepsilon^{\alpha\lambda\mu} \alpha^m F_{\alpha\mu}^n) \approx 0.$$

In the Chern–Simons model, the total conserved current T_{CS} is equal to zero. At the same time, if we add the *Chern–Simons Lagrangian density* to the Yang–Mills one, T_{CS} plays the role of the massive term and makes the contribution into the standard Noether current of the Yang–Mills gauge theory.

Let τ be a vector-field on the base X and τ_B its lift onto the bundle C by means of a section B of C . Remind that the vector-fields τ_B are the general principal vector-fields associated with local 1-parameter groups of general gauge isomorphisms of C . We calculate

$$\mathfrak{L}_{\tau_B} L_{CS} = \frac{1}{k} a_{mn}^G \varepsilon^{\alpha\lambda\mu} \partial_\alpha (\tau^\nu B_\nu^m \partial_\alpha A_\mu^n) d^3x.$$

The corresponding SEM transformation law takes the form

$$\begin{aligned} d_\alpha \mathcal{J}_{CS}^\alpha &= d_\alpha (\mathcal{J}^\alpha - \frac{1}{2k} a_{mn}^G \varepsilon^{\alpha\lambda\mu} \tau^\nu B_\nu^m \partial_\alpha A_\mu^n) \approx 0, \quad \text{where} \quad (5.382) \\ \mathcal{J}^\alpha &= \pi_n^{\mu\lambda} [\tau^\nu \partial_\nu A_\mu^m - \tau^\nu (\partial_\mu B_\nu^n - c_{pq}^n A_\mu^p B_\nu^q) - \partial_\mu \tau^\nu (B_\nu^n - A_\nu^n)] - \delta_\nu^\alpha \tau^\nu \mathcal{L}_{CS} \end{aligned}$$

is the standard SEM-tensor relative to the lift τ_B of the vector-field τ .

Let A be a critical section. We consider the lift of the vector-field τ on X onto C by means of the principal connection $B = A$. Then, the energy-momentum conservation law (5.382) becomes the conservation law

$$d_\alpha \left[\frac{1}{k} a_{mn}^G \tau^\nu \varepsilon^{\alpha\lambda\mu} A_\alpha^m \mathcal{F}_{\nu\mu}^n - \tau^\alpha \mathcal{L}_{CS} \right] = d_\alpha \left(-\frac{1}{6k} \tau^\alpha \varepsilon^{\alpha\nu\mu} c_{npq} A_\alpha^n A_\nu^p A_\mu^q \right) \approx 0. \quad (5.383)$$

Note that, since the gauge symmetry of the Chern–Simons Lagrangian density is broken, the energy-momentum conservation law (5.383) fails to be invariant under gauge transformations.

Let us consider Lagrangian densities of topological gauge models which are invariant under the general gauge isomorphisms of the bundle C . Though they imply the zero Euler–Lagrangian operators, the corresponding strong identities may be used as the superpotential terms when such a topological Lagrangian density is added to the Yang–Mills one.

Let $P \rightarrow X$ be a principal bundle with the structure Lie group G . Let us consider the bundle $J^1(Q, P) \rightarrow C$. This also is a G principal bundle. Due to the canonical vertical splitting $VP = P \times \mathfrak{g}_l$, where \mathfrak{g}_l is the left Lie algebra of the group G , the complementary map (5.6) of $J^1(Q, P)$ defines the canonical \mathcal{G} -valued one-form θ on $J^1(Q, P)$. This form is the connection form of the canonical principal connection on the principal bundle $J^1(Q, P) \rightarrow C$ [García (1972)]. Moreover, if $\Gamma_A : P \rightarrow J^1(Q, P)$ is

a principal connection on P and A the corresponding connection form, we have $\Gamma_A^* \theta = A$. If Ω and R_A are the curvature 2-forms of the connections θ and A respectively, then $\Gamma_A^* \Omega = R_A$.

Local connection 1-forms on C associated with the canonical connection θ are given by the coordinate expressions $k_\mu^m dx^\mu \otimes I_m$. The corresponding curvature two-form on C reads

$$\Omega_C = (dk_\mu^m \wedge dx^\mu - \frac{1}{2} c_{nl}^m k_\mu^n k_\nu^l dx^\nu \wedge dx^\mu) \otimes I_m.$$

Let $I(\mathfrak{g})$ be the algebra of real G -invariant polynomial on the Lie algebra \mathfrak{g} of the group G . Then, there is the well-known *Weyl homomorphism* of $I(\mathfrak{g})$ into the *de Rham cohomology algebra* $H^*(C, \mathbb{R})$. Using this isomorphism, every k -linear element $r \in I(\mathfrak{g})$ is represented by the cohomology class of the closed characteristic $2k$ -form $r(\Omega_C)$ on C . If A is a section of C , we have $A^* r(\Omega_C) = r(F)$, where F is the strength of A and $r(F)$ is the corresponding characteristic form on X .

Let $\dim X$ be even and a characteristic n -form $r(\Omega_C)$ on C exist. This is a Lepagean form which defines a gauge-invariant Lagrangian density

$$L_r = h_0(r(\Omega_C))$$

on the jet space $J^1(Q, C)$. The Euler-Lagrangian operator associated with L_r is equal to zero. Then, for any projectable vector-field u on C , we have the strong relation (5.251):

$$\mathfrak{L}_{\bar{u}} h_0(r(\Omega_C)) = h_0(d\bar{u} \rfloor r(\Omega_C)).$$

If u is a general principal vector-field on C , this relation takes the form

$$d_H(\bar{u} \rfloor r(\Omega_C)) = 0.$$

For example, let $\dim X = 4$ and the group G be semisimple. Then, the characteristic Chern-Pontryagin 4-form

$$r(\Omega_C) = a_{mn}^G \Omega_C^n \wedge \Omega_C^m.$$

It is the Lepagean equivalent of the *Chern-Pontryagin Lagrangian density*

$$L = \frac{1}{k} a_{mn}^G \varepsilon^{\alpha\beta\mu\nu} \mathcal{F}_{\alpha\beta}^n \mathcal{F}_{\mu\nu}^m d^4 x$$

of the topological Yang-Mills theory.

5.12 Application: Modern Geometrodynamics

In this subsection we present some modern developments of the classical Einstein–Wheeler geometrodynamics that we briefly reviewed as a motivation to our geometrical machinery.

5.12.1 Stress–Energy–Momentum Tensors

While in analytical mechanics there exists the conventional differential energy conservation law, in field theory it does not exist (see [Sardanashvily (1998)]). Let F be a smooth manifold. In time-dependent mechanics on the phase-space $\mathbb{R} \times T^*F$ coordinated by (t, y^i, \dot{y}_i) and on the configuration space $\mathbb{R} \times TF$ coordinated by (t, y^i, \dot{y}^i) , the Lagrangian energy and the construction of the Hamiltonian formalism require the prior choice of a connection on the bundle $\mathbb{R} \times F \rightarrow \mathbb{R}$. However, such a connection is usually hidden by using the natural trivial connection on this bundle. Therefore, given a Hamiltonian function \mathcal{H} on the phase-space manifold $\mathbb{R} \times T^*F$, we have the usual *energy conservation law*

$$\frac{d\mathcal{H}}{dt} \approx \frac{\partial \mathcal{H}}{\partial t} \quad (5.384)$$

where by ‘ \approx ’ is meant the weak identity modulo the Hamiltonian equations. Given a Lagrangian function \mathcal{L} on the configuration manifold $\mathbb{R} \times TF$, there exists the fundamental identity

$$\frac{\partial \mathcal{L}}{\partial t} + \frac{d}{dt}(y^i(t) \frac{\partial \mathcal{L}}{\partial \dot{y}^i} - \mathcal{L}) \approx 0 \quad (5.385)$$

modulo the equations of motion. It is the *energy conservation law* in the following sense. Let \widehat{L} be the Legendre morphism given by $\dot{y}_i \circ \widehat{L} = \partial_{\dot{y}_i} \mathcal{L}$, and $Q = \text{Im } \widehat{L}$ the Lagrangian constraint manifold. Let \mathcal{H} be a Hamiltonian function associated with \mathcal{L} and \widehat{H} the momentum morphism, $\dot{y}^i \circ \widehat{H} = \partial_{\dot{y}_i} \mathcal{H}$. Every solution r of the Hamiltonian equations of \mathcal{H} which lives on Q yields the solution $\widehat{H} \circ r$ of the Euler–Lagrangian equations of \mathcal{L} . Then, the identity (5.385) on $\widehat{H} \circ r$ recovers the energy transformation law (5.384) on r .⁴

⁴There are different Hamiltonian functions associated with the same singular Lagrangian function as a rule. Given such a Hamiltonian function, the Lagrangian constraint space Q plays the role of the primary constraint space, and the Dirac procedure can be used in order to get the final constraint space where a solution of the Hamiltonian equations exists [Gotay *et. al.* (1978); León and Marrero (1993)].

Recall that in field theory, classical fields are described by sections of a fibre bundle $Y \rightarrow X$, while their dynamics is phrased in terms of jet spaces. We restrict ourselves to the *first-order Lagrangian formalism* when the configuration space is $J^1(X, Y)$. Given fibred coordinates (x^μ, y^i) of Y , the jet space $J^1(X, Y)$ is equipped with the adapted coordinates (x^μ, y^i, y_μ^i) . Recall that the first-order Lagrangian density on $J^1(X, Y)$ is defined to be the morphism

$$\begin{aligned} L : J^1(X, Y) &\rightarrow \wedge^n T^*X, & (n = \dim X), \\ L &= \mathcal{L}(x^\mu, y^i, y_\mu^i)\omega, & \text{with } \omega = dx^1 \wedge \dots \wedge dx^n, \end{aligned}$$

while the corresponding first-order Euler–Lagrangian equations for sections \bar{s} of the jet bundle $J^1(X, Y) \rightarrow X$ read

$$\partial_\alpha \bar{s}^i = \bar{s}_\alpha^i, \quad \partial_i \mathcal{L} - (\partial_\alpha + \bar{s}_\alpha^j \partial_j + \partial_\alpha \bar{s}_\mu^j \partial_j^\mu) \partial_i^\alpha \mathcal{L} = 0. \quad (5.386)$$

As before, we consider the *Lie derivatives of Lagrangian densities* in order to get *differential conservation laws*. Let $u = u^\mu(x) \partial_\mu + u^i(y) \partial_i$ be a projectable vector-field on $Y \rightarrow X$ and \bar{u} its jet lift (5.9) onto $J^1(X, Y) \rightarrow X$. Given L , let us calculate the Lie derivative $\mathfrak{L}_{\bar{u}} L$. We get the identity

$$\bar{s}^* \mathfrak{L}_{\bar{u}} L \approx -\frac{d}{dx^\alpha} [\pi_i^\alpha (u^\mu \bar{s}_\mu^i - u^i) - u^\alpha \mathcal{L}] \omega, \quad \text{with } \pi_i^\mu = \partial_i^\mu \mathcal{L}, \quad (5.387)$$

modulo the Euler–Lagrangian equations (5.386). In particular, if u is a vertical vector-field this identity becomes the current conservation law exemplified by the Noether identities in gauge theory [Sardanashvily (1994)].

Let now $\tau = \tau^\alpha \partial_\alpha$ be a vector-field on X and

$$\tau_\Gamma = \tau^\mu (\partial_\mu + \Gamma_\mu^i \partial_i) \quad (5.388)$$

its horizontal lift onto Y by a connection Γ on $Y \rightarrow X$. In this case, the identity (5.387) takes the form

$$\bar{s}^* \mathfrak{L}_{\bar{\tau}_\Gamma} L \approx -\frac{d}{dx^\alpha} [\tau^\mu \mathcal{J}_\Gamma^\alpha{}_\mu(\bar{s})] \omega, \quad (5.389)$$

$$\text{where } \mathcal{J}_\Gamma^\alpha{}_\mu(\bar{s}) = [\pi_i^\alpha (y_\mu^i - \Gamma_\mu^i) - \delta_\mu^\alpha \mathcal{L}] \circ \bar{s}_\mu^i$$

is the stress–energy–momentum (SEM) tensor on a field \bar{s} relative to the connection Γ . We here restrict ourselves to this particular case of SEM-tensors [Kijowski and Tulczyjew (1979)].

For example, let us choose the trivial local connection $\Gamma_\mu^i = 0$. In this case, the identity (5.389) recovers the well-known conservation law

$$\frac{\partial \mathcal{L}}{\partial x^\alpha} + \frac{d}{dx^\alpha} \mathcal{J}^\alpha{}_\mu(\bar{s}) \approx 0$$

of the canonical energy-momentum tensor

$$\mathcal{J}^\alpha{}_\mu(\bar{s}) = \pi_i^\alpha \bar{s}_\mu^i - \delta_\mu^\alpha \mathcal{L}. \quad (5.390)$$

Physicists often lose sight of the fact that (5.390) fails to be a mathematical well-behaved object. The crucial point lies in the fact that the Lie derivative

$$\mathfrak{L}_{\bar{\tau}} L = \{\partial_\mu \tau^\mu \mathcal{L} + [\tau^\mu \partial_\mu + \tau^\mu \Gamma_\mu^i \partial_i + (\partial_\alpha (\tau^\mu \Gamma_\mu^i) + \tau^\mu y_\alpha^j \partial_j \Gamma_\mu^i - y_\mu^i \partial_\alpha \tau^\mu) \partial_i] \mathcal{L}\}_\omega$$

is almost never equal to zero. Therefore, it is not obvious how to choose the true energy-momentum tensor.

The canonical energy-momentum tensor (5.390) in gauge theory is symmetrized by hand in order to get the gauge-invariant one. In gauge theory in the presence of a background world metric g , the identity (5.389) is brought into the covariant conservation law for the *metric SEM-tensor*,

$$\nabla_\alpha t^\alpha{}_\mu \approx 0. \quad (5.391)$$

In Einstein's general relativity, the covariant conservation law (5.391) issues directly from the gravitation equations. It is concerned with the zero-spin matter in the presence of the gravitational field generated by this matter, though the matter is not required to satisfy the motion equations. The total energy-momentum conservation law for matter and gravity is introduced by hand. It is usually written as

$$\frac{d}{dx^\mu} [(-g)^N (t^{\lambda\mu} + t_g^{\lambda\mu})] \approx 0, \quad (5.392)$$

where the energy-momentum pseudotensor $t_g^{\lambda\mu}$ of a metric gravitational field is defined to satisfy the relation

$$(-g)^N (t^{\lambda\mu} + t_g^{\lambda\mu}) \approx \frac{1}{2\kappa} \partial_\sigma \partial_\alpha [(-g)^N (g^{\lambda\mu} g^{\sigma\alpha} - g^{\sigma\mu} g^{\lambda\alpha})]$$

modulo the Einstein equations. However, the conservation law (5.392) appears satisfactory only in cases of asymptotic-flat gravitational fields and of a background metric.

Moreover, the covariant conservation law (5.391) fails to take place in the affine-metric gravitation theory and in the gauge gravitation theory, e.g., in the presence of fermion fields.

Thus, we have not any conventional energy-momentum conservation law in Lagrangian field theory. In particular, one may take different SEM-tensors for different field models and, moreover, different SEM-tensors for different solutions of the same field equations. Just the latter in fact is the above-mentioned symmetrization of the canonical energy-momentum tensor in gauge theory.

Gauge theory exemplifies constraint field theories. Contemporary field models are almost always the constraint ones. To describe them, let us turn to the Hamiltonian formalism.

When applied to field theory, the conventional Hamiltonian formalism takes the form of the instantaneous Hamiltonian formalism where canonical variables are field functions at a given instant of time. The corresponding phase-space is infinite-dimensional, so that the Hamiltonian equations in the bracket form fail to be differential equations.

The true partners of the Lagrangian formalism in classical field theory are polysymplectic and multisymplectic Hamiltonian machineries where canonical momenta correspond to derivatives of fields with respect to all world coordinates, not only the temporal one [Cariñena *et. al.* (1991); Sardanashvily (1993)]. We here follow the multimomentum Hamiltonian formulation of field theory when the phase-space of fields is the Legendre bundle over Y

$$\Pi = \wedge^n T^*X \otimes TX \otimes V^*Y, \quad (5.393)$$

which is coordinated by $(x^\alpha, y^i, p_i^\alpha)$ [Sardanashvily (1993); Sardanashvily (1994)]. Every Lagrangian density L on $J^1(X, Y)$ implies the Legendre morphism

$$\widehat{L}: J^1(X, Y) \rightarrow \Pi, \quad p_i^\mu \circ \widehat{L} = \pi_i^\mu.$$

The Legendre bundle (5.393) carries the polysymplectic form

$$\Omega = dp_i^\alpha \wedge dy^i \wedge \omega \otimes \partial_\alpha. \quad (5.394)$$

Recall that one says that a connection γ on the fibred Legendre manifold $\Pi \rightarrow X$ is a *Hamiltonian connection* if the form $\gamma \rfloor \Omega$ is closed. Then, a Hamiltonian form H on Π is defined to be an exterior form such that

$$dH = \gamma \rfloor \Omega \quad (5.395)$$

for some Hamiltonian connection γ . The key point lies in the fact that every Hamiltonian form admits the following splitting

$$H = p_i^\alpha dy^i \wedge \omega_\alpha - p_i^\alpha \Gamma_\alpha^i \omega - \tilde{\mathcal{H}}_\Gamma \omega = p_i^\alpha dy^i \wedge \omega_\alpha - \mathcal{H} \omega, \quad \omega_\alpha = \partial_\alpha \rfloor \omega, \quad (5.396)$$

where Γ is a connection on $Y \rightarrow X$.

Given the splitting (5.396), the equality (5.395) becomes the Hamiltonian equations

$$\partial_\alpha r^i = \partial_\alpha^i \mathcal{H}, \quad \partial_\alpha r_i^\alpha = -\partial_i \mathcal{H} \quad (5.397)$$

for sections r of $\Pi \rightarrow X$.

The Hamiltonian equations (5.397) are the multimomentum generalization of the standard Hamiltonian equations in mechanics. The corresponding multimomentum generalization of the conventional energy conservation law (5.384) is the weak identity

$$\tau^\mu [(\partial_\mu + \Gamma_\mu^i \partial_i - \partial_i \Gamma_\mu^j r_j^\alpha \partial_\alpha^i) \tilde{\mathcal{H}}_\Gamma - \frac{d}{dx^\alpha} T_\Gamma^\alpha{}_\mu(r)] \approx \tau^\mu r_i^\alpha R_{\lambda\mu}^i, \quad (5.398)$$

$$T_\Gamma^\alpha{}_\mu(r) = [r_i^\alpha \partial_\mu^i \tilde{\mathcal{H}}_\Gamma - \delta_\mu^\alpha (r_i^\alpha \partial_\alpha^i \tilde{\mathcal{H}}_\Gamma - \tilde{\mathcal{H}}_\Gamma)], \quad (5.399)$$

$$\text{where} \quad R_{\lambda\mu}^i = \partial_\alpha \Gamma_\mu^i - \partial_\mu \Gamma_\alpha^i + \Gamma_\alpha^j \partial_j \Gamma_\mu^i - \Gamma_\mu^j \partial_j \Gamma_\alpha^i$$

is the curvature of the connection Γ . One can think of the tensor (5.399) as being the Hamiltonian SEM-tensor.

If a Lagrangian density is regular, the multimomentum Hamiltonian formalism is equivalent to the Lagrangian formalism, otherwise in case of degenerate Lagrangian densities. In field theory, if a Lagrangian density is not regular, the Euler-Lagrangian equations become underdetermined and require supplementary gauge-type conditions. In gauge theory, they are the familiar gauge conditions. However, in general case, the gauge-type conditions remain elusive. In the framework of the multimomentum Hamiltonian formalism, they appear automatically as a part of the Hamiltonian equations. The key point consists in the fact that, given a degenerate Lagrangian density, one must consider a family of different associated Hamiltonian forms in order to exhaust all solutions of the Euler-Lagrangian equations [Cariñena *et al.* (1991); Sardanashvily (1993)].

Lagrangian densities of all realistic field models are quadratic or affine in the velocity coordinates y_μ^i . Complete family of Hamiltonian forms associated with such a Lagrangian density always exists [Sardanashvily (1993); Sardanashvily (1994)]. Moreover, these Hamiltonian forms differ from each

other only in connections Γ in the splitting (5.396). Different connections are responsible for different gauge-type conditions mentioned above. They are also the connections which one should use in construction of the Hamiltonian SEM-tensors (5.399).

The identity (5.398) remains true in the first-order Lagrangian theories of gravity. In this work, we examine the metric-affine gravity where independent dynamical variables are world metrics and general linear connections. The energy-momentum conservation law in the affine-metric gravitation theory is not widely discussed. We construct the Hamiltonian SEM-tensor for gravity. In case of the affine Hilbert-Einstein Lagrangian density, it is equal to

$$T_{\mu}^{\alpha} = \frac{1}{2\kappa} \delta_{\mu}^{\alpha} R \sqrt{-g}$$

and the total conservation law (5.398) for matter and gravity is reduced to the conservation law for matter in the presence of a background world metric, otherwise in case of quadratic Lagrangian densities.

Lagrangian SEM-Tensors

Given a Lagrangian density L , the jet space $J^1(X, Y)$ carries the associated Poincaré-Cartan form [Sardanashvily (1998)]

$$\Xi_L = \pi_i^{\alpha} dy^i \wedge \omega_{\alpha} - \pi_i^{\alpha} y_{\alpha}^i \omega + \mathcal{L}\omega \quad (5.400)$$

and the Lagrangian polysymplectic form

$$\Omega_L = (\partial_j \pi_i^{\alpha} dy^j + \partial_j^{\mu} \pi_i^{\alpha} dy_{\mu}^j) \wedge dy^i \wedge \omega \otimes \partial_{\alpha}.$$

Using the pull-back of these forms onto the repeated jet space $J^1 J^1(X, Y)$, one can construct the exterior generating form on $J^1 J^1(X, Y)$,

$$\begin{aligned} \Lambda_L &= d\Xi_L - \lambda] \Omega_L = [y_{(\lambda)}^i - y_{\alpha}^i] d\pi_i^{\alpha} + (\partial_i - \widehat{\partial}_{\alpha} \partial_i^{\alpha}) \mathcal{L} dy^i \wedge \omega, \quad (5.401) \\ \lambda &= dx^{\alpha} \otimes \widehat{\partial}_{\alpha}, \quad \widehat{\partial}_{\alpha} = \partial_{\alpha} + y_{(\lambda)}^i \partial_i + y_{\mu\lambda}^i \partial_i^{\mu}. \end{aligned}$$

Its restriction to the sesquiholonomic jet space $\widehat{J}^2(X, Y)$ defines the first-order Euler-Lagrangian operator

$$\begin{aligned} \mathcal{E}'_L : \widehat{J}^2(X, Y) &\longrightarrow \wedge^{n+1} T^*Y, \quad \text{given by} \\ \mathcal{E}'_L &= \delta_i \mathcal{L} dy^i \wedge \omega = [\partial_i - (\partial_{\alpha} + y_{\alpha}^i \partial_i + y_{\mu\lambda}^i \partial_i^{\mu}) \partial_i^{\alpha}] \mathcal{L} dy^i \wedge \omega, \quad (5.402) \end{aligned}$$

corresponding to L . The restriction of the form (5.401) to the second-order jet space $J^2(X, Y)$ of Y recovers the familiar variational Euler–Lagrangian operator

$$\mathcal{E}_L : J^2(X, Y) \longrightarrow \wedge^{n+1} T^*Y,$$

given by the expression (5.402), but with symmetric coordinates $y_{\mu\lambda}^i = y_{\lambda\mu}^i$.

Let \bar{s} be a section of the jet bundle $J^1(X, Y) \rightarrow X$ such that its jet prolongation $j^1\bar{s}$ takes its values into $\text{Ker } \mathcal{E}'_L$ given by the coordinate relations

$$\partial_i \mathcal{L} - (\partial_\alpha + y_\alpha^j \partial_j + y_{\mu\lambda}^j \partial_j^\mu) \partial_i^\alpha \mathcal{L} = 0.$$

Then, \bar{s} satisfies the first-order Euler–Lagrangian equations (5.386). These equations are equivalent to the second-order Euler–Lagrangian equations

$$\partial_i \mathcal{L} - (\partial_\alpha + \partial_\alpha s^j \partial_j + \partial_\alpha \partial_\mu s^j \partial_j^\mu) \partial_i^\alpha \mathcal{L} = 0. \quad (5.403)$$

for sections s of $Y \rightarrow X$ where $\bar{s} = j^1 s$.

We have the following differential conservation laws on solutions of the first-order Euler–Lagrangian equations.

Given a Lagrangian density L on $J^1(X, Y)$, let us consider its pull-back onto $\widehat{J}^2(X, Y)$. Let u be a projectable vector-field on $Y \rightarrow X$ and \bar{u} its jet lift (5.9) onto $J^1(X, Y) \rightarrow X$. Its pull-back onto $J^1 J^1(X, Y)$ has the canonical horizontal splitting (5.14) given by the expression

$$\bar{u} = \bar{u}_H + \bar{u}_V = u^\alpha (\partial_\alpha + y_{(\lambda}^i \partial_i + y_{\mu\lambda}^i \partial_i^\mu) + [(u^i - y_{(\lambda}^i) u^\alpha) \partial_i + (u_\mu^i - y_{\mu\lambda}^i u^\alpha) \partial_i^\mu].$$

Let us calculate the Lie derivative $\mathfrak{L}_{\bar{u}} L$. We have

$$\begin{aligned} \mathfrak{L}_{\bar{u}} L &= \widehat{\partial}_\alpha [\pi_i^\lambda (u^i - u^\mu y_\mu^i) + u^\alpha \mathcal{L}] \omega + \bar{u}_V] \mathcal{E}'_L, \\ \widehat{\partial}_\alpha &= \partial_\alpha + y_\alpha^i \partial_i + y_{\mu\lambda}^i \partial_i^\mu. \end{aligned} \quad (5.404)$$

Being restricted to $\text{Ker } \mathcal{E}'_L$, the equality (5.404) is written

$$\begin{aligned} &\partial_\alpha u^\alpha \mathcal{L} + [u^\alpha \partial_\alpha + u^i \partial_i + (\partial_\alpha u^i + y_\alpha^j \partial_j u^i - y_\mu^i \partial_\alpha u^\mu) \partial_i^\alpha] \mathcal{L} \\ &\approx \widehat{\partial}_\alpha [\pi_i^\alpha (u^i - u^\mu y_\mu^i) + u^\alpha \mathcal{L}]. \end{aligned} \quad (5.405)$$

On solutions \bar{s} of the first-order Euler–Lagrangian equations, the weak identity (5.405) becomes the differential conservation law

$$\bar{s}^* \mathfrak{L}_{\bar{u}} L \approx d(\bar{u}] \Xi_L \circ \bar{s}),$$

which takes the coordinate form (5.387).

In particular, let τ_Γ be the horizontal lift (5.388) of a vector-field τ on X onto $Y \rightarrow X$ by a connection Γ on Y . In this case, the identity (5.405) is written

$$\tau^\mu \{ [\partial_\mu + \Gamma_\mu^i \partial_i + (\partial_\alpha \Gamma_\mu^i + y_\alpha^j \partial_j \Gamma_\mu^i) \partial_i^\alpha] \mathcal{L} + \hat{\partial}_\alpha [\pi_i^\alpha (y_\mu^i - \Gamma_\mu^i) - \delta_\mu^\alpha \mathcal{L}] \} \approx 0. \quad (5.406)$$

On solutions \bar{s} of the first-order Euler-Lagrangian equations, the identity (5.406) becomes the differential conservation law (5.389) where $\mathcal{J}_\Gamma^\alpha{}_\mu(\bar{s})$ are coefficients of the T^*X -valued form on X ,

$$\mathcal{J}_\Gamma(\bar{s}) = -(\Gamma] \Xi_L) \circ \bar{s} = [\pi_i^\alpha (\bar{s}_\mu^i - \Gamma_\mu^i) - \delta_\mu^\alpha \mathcal{L}] dx^\mu \otimes \omega_\alpha. \quad (5.407)$$

This conservation law takes the coordinate form

$$\tau^\mu \{ [\partial_\mu + \Gamma_\mu^i \partial_i + (\partial_\alpha \Gamma_\mu^i + \bar{s}_\alpha^j \partial_j \Gamma_\mu^i) \partial_i^\alpha] \mathcal{L} + \frac{d}{dx^\alpha} [\pi_i^\alpha (\bar{s}_\mu^i - \Gamma_\mu^i) - \delta_\mu^\alpha \mathcal{L}] \} \approx 0.$$

SEM Conservation Laws

Every projectable vector-field u on the bundle $Y \rightarrow X$ which covers a vector-field τ on the base X is represented as the sum of a vertical vector-field on $Y \rightarrow X$ and some lift of τ onto Y . Hence, any differential transformation law (5.249) can be represented as a superposition of some transformation law associated with a vertical vector-field on the bundle $Y \rightarrow X$ and the one induced by the lift of a vector-field on the base X onto Y . Therefore, we can reduce our consideration to transformation laws associated with these two types of vector-fields on Y .

Vertical vector-fields result in transformation and conservation laws of Noether currents. In general case, a vector-field τ on a base X induces a vector-field on Y only by means of some connection on the bundle $Y \rightarrow X$. Such lifts result in the transformation laws of the SEM-tensors.

Given a bundle $Y \rightarrow X$, let τ be a vector-field on X and

$$\tau_\Gamma = \tau] \Gamma = \tau^\mu (\partial_\mu + \Gamma_\mu^i \partial_i)$$

its horizontal lift onto $Y \rightarrow X$ by means of a connection on Y , given by

$$\Gamma = dx^\mu \otimes (\partial_\mu + \Gamma_\mu^i \partial_i).$$

In this case, the weak identity (5.405) is written

$$\begin{aligned} & \partial_\mu \tau^\mu \mathcal{L} + [\tau^\mu \partial_\mu + \tau^\mu \Gamma_\mu^i \partial_i + (\partial_\alpha (\tau^\mu \Gamma_\mu^i) + \tau^\mu y_\alpha^j \partial_j \Gamma_\mu^i - y_\mu^i \partial_\alpha \tau^\mu) \partial_i^\alpha] \mathcal{L} \\ & - \hat{\partial}_\alpha [\pi_i^\alpha (\tau^\mu \Gamma_\mu^i - \tau^\mu y_\mu^i) + \delta_\mu^\alpha \tau^\mu \mathcal{L}] \approx 0. \end{aligned} \quad (5.408)$$

One can simplify it as follows:

$$\tau^\mu \{ [\partial_\mu + \Gamma_\mu^i \partial_i + (\partial_\alpha \Gamma_\mu^i + y_\alpha^j \partial_j \Gamma_\mu^i) \partial_i^\alpha] \mathcal{L} - \widehat{\partial}_\alpha [\pi_i^\alpha (\Gamma_\mu^i - y_\mu^i) + \delta_\mu^\alpha \mathcal{L}] \approx 0.$$

Let us emphasize that this relation takes place for arbitrary vector-field τ on X . Therefore, it is equivalent to the system of the weak identities

$$[\partial_\mu + \Gamma_\mu^i \partial_i + (\partial_\alpha \Gamma_\mu^i + y_\alpha^j \partial_j \Gamma_\mu^i) \partial_i^\alpha] \mathcal{L} - \widehat{\partial}_\alpha [\pi_i^\alpha (\Gamma_\mu^i - y_\mu^i) + \delta_\mu^\alpha \mathcal{L}] \approx 0. \quad (5.409)$$

On solutions s of the Euler-Lagrangian equations, the weak identity (5.408) becomes the weak transformation law

$$s^* \mathfrak{L}_{\bar{\tau}_\Gamma} L + \frac{d}{dx^\alpha} [\tau^\mu \mathcal{J}_{\Gamma^\alpha \mu}(s)] \omega \approx 0$$

and to the equivalent system of the weak transformation laws

$$[\partial_\mu + \Gamma_\mu^i \partial_i + (\partial_\alpha \Gamma_\mu^i + \partial_\alpha s^j \partial_j \Gamma_\mu^i) \partial_i^\alpha] \mathcal{L} + \frac{d}{dx^\alpha} [\pi_i^\alpha (\partial_\mu s^i - \Gamma_\mu^i) - \delta_\mu^\alpha \mathcal{L}] \approx 0 \quad (5.410)$$

where $\mathcal{J}_{\Gamma^\alpha \mu}(s)$ is the SEM-tensor given by the components of the T^*X -valued $(n-1)$ -form on X ,

$$\mathcal{J}_\Gamma(s) = -(\Gamma] \Xi_L) \circ s = [\pi_i^\alpha (\partial_\mu s^i - \Gamma_\mu^i) - \delta_\mu^\alpha \mathcal{L}] dx^\mu \otimes \omega_\alpha.$$

It is clear that the first and the second terms in (5.410) taken separately fail to be well-behaved objects. Therefore, only their combination may result in the satisfactory transformation or conservation law.

For example, let a Lagrangian density L depend on a background metric g on the base X . In this case, we have

$$\partial_\mu \mathcal{L} = -t_\beta^\alpha \sqrt{|g|} \Gamma_{\mu\alpha}^\beta, \quad \text{where} \quad t_\beta^\alpha = g^{\alpha\gamma} t_{\gamma\beta}$$

is the *metric SEM-tensor* (by definition), while $\Gamma_{\mu\alpha}^\beta$ are the *Christoffel symbols* of the metric g . Then, the weak transformation law (5.410) takes the form

$$\begin{aligned} & -t_\beta^\alpha \sqrt{|g|} \Gamma_{\mu\alpha}^\beta + [\Gamma_\mu^i \partial_i + (\partial_\alpha \Gamma_\mu^i + \partial_\alpha s^j \partial_j \Gamma_\mu^i) \partial_i^\alpha] \mathcal{L} \\ & + \frac{d}{dx^\alpha} [\pi_i^\alpha (\partial_\mu s^i - \Gamma_\mu^i) - \delta_\mu^\alpha \mathcal{L}] \approx 0, \end{aligned}$$

and, under suitable conditions of symmetries of the Lagrangian density L , it may become the covariant conservation law $\nabla_\alpha t_\beta^\alpha = 0$ where ∇_α denotes the covariant derivative relative to the connection $\Gamma_{\mu\alpha}^\beta$.

Note that, if we consider another Lepagean equivalent of the Lagrangian density L , the SEM transformation law takes the form

$$s^* \mathfrak{L}_{\bar{\tau}_\Gamma} L + \frac{d}{dx^\alpha} [\tau^\mu \mathcal{J}'_{\Gamma^\alpha \mu}(s)] \omega \approx 0,$$

$$\text{where} \quad \mathcal{J}'_{\Gamma^\alpha \mu} = \mathcal{J}_{\Gamma^\alpha \mu} - \frac{d}{dx^\nu} [(\partial_\mu s^i - \Gamma_\mu^i) c_i^{\lambda\nu}],$$

that is, the SEM-tensors $\mathcal{J}'_{\Gamma^\alpha \mu}$ and $\mathcal{J}_{\Gamma^\alpha \mu}$ differ from each other in the superpotential-type term: $-\frac{d}{dx^\nu} [(\partial_\mu s^i - \Gamma_\mu^i) c_i^{\lambda\nu}]$.

In particular, if the bundle Y has a fibre metric a_{ij}^Y , one can choose

$$c_i^{\mu\nu} = a_{ij}^Y g^{\mu\alpha} g^{\nu\beta} R_{\alpha\beta}^j,$$

where R is the curvature of the connection Γ on the bundle Y and g is a metric on X . In this case, the superpotential contribution into the SEM-tensor is equal to $-\frac{d}{dx^\nu} [a_{ij}^Y g^{\lambda\alpha} g^{\nu\beta} (\partial_\mu s^i - \Gamma_\mu^i) R_{\alpha\beta}^j]$.

Let us now consider the weak identity (5.408) when a vector-field τ on the base X induces a vector-field on Y by means of different connections Γ and Γ' on $Y \rightarrow X$. Their difference result in the weak identity

$$[\tau^\mu \sigma_\mu^i \partial_i + (\partial_\alpha (\tau^\mu \sigma_\mu^i) + y_\alpha^j \partial_j (\tau^\mu \sigma_\mu^i)) \partial_i^\alpha] \mathcal{L} - \widehat{\partial}_\alpha [\pi_i^\alpha \tau^\mu \sigma_\mu^i] \approx 0 \quad (5.411)$$

where $\sigma = \Gamma' - \Gamma$ is a soldering form on the bundle $Y \rightarrow X$ and

$$\tau \rfloor \sigma = \tau^\mu \sigma_\mu^i \partial_i \quad (5.412)$$

is a vertical vector-field. It is clear that the identity (5.411) is exactly the weak identity (5.405) in case of the vertical vector-field (5.412).

It follows that every SEM transformation law contains a Noether transformation law. Conversely, every Noether transformation law associated with a vertical vector-field u_V on $Y \rightarrow X$ can be get as the difference of two SEM transformation laws if the vector-field u_V takes the form $u_V = \tau \rfloor \sigma$, where σ is some soldering form on Y and τ is a vector-field on X . In field theory, this representation fails to be unique. On the contrary, in Newtonian mechanics there is the 1-1 correspondence between the vertical vector-fields and the soldering forms on the bundle $\mathbb{R} \times F \rightarrow F$.

Note that one can consider the pull-back of the first-order Lagrangian density L and their Lepagean equivalents onto the infinite order jet space $J^\infty Y$. In this case, there exists the canonical lift τ_H^∞ (5.59) of a vector-field τ on X onto $J^\infty Y$. One can treat this lift as the horizontal lift of τ by

means of the canonical connection on the bundle $J^\infty Y \rightarrow X$, given by

$$\Gamma_\infty = dx^\mu \otimes (\partial_\mu + y^i \partial_i + y_\alpha^i \partial_i^\alpha + \cdots).$$

Multimomentum Hamiltonian Formalism

Let Π be the Legendre bundle (5.393) coordinated by $(x^\alpha, y^i, p_i^\alpha)$. By $J^1(X, \Pi)$ is meant the first-order jet space of $\Pi \rightarrow X$. It is coordinated by $(x^\alpha, y^i, p_i^\alpha, y_{(\mu)}^i, p_{i\mu}^\alpha)$. The Legendre manifold Π carries the generalized Liouville form

$$\theta = -p_i^\alpha dy^i \wedge \omega \otimes \partial_\alpha$$

and the polysymplectic form Ω (5.394).

The Hamiltonian formalism in fibred manifolds is formulated intrinsically in terms of Hamiltonian connections which play the role similar to that of Hamiltonian vector-fields in the symplectic geometry [Sardanashvily (1993)].

We say that a jet field (resp. a connection)

$$\gamma = dx^\alpha \otimes (\partial_\alpha + \gamma_{(\lambda)}^i \partial_i + \gamma_{i\lambda}^\mu \partial_\mu^i)$$

on the Legendre manifold $\Pi \rightarrow X$ is a Hamiltonian jet field (resp. a Hamiltonian connection) if the following exterior form is closed:

$$\gamma \rfloor \Omega = dp_i^\alpha \wedge dy^i \wedge \omega_\alpha + \gamma_{i\lambda}^\alpha dy^i \wedge \omega - \gamma_{(\lambda)}^i dp_i^\alpha \wedge \omega.$$

An exterior n -form H on the Legendre manifold Π is called a *Hamiltonian form* if, on an open neighborhood of each point of Π , there exists a *Hamiltonian jet-field* satisfying the equation $\gamma \rfloor \Omega = dH$, i.e., if there exists a Hamiltonian connection satisfying the equation (5.395). Hamiltonian connections constitute an affine subspace of connections on $\Pi \rightarrow X$. The following construction shows that this subspace is not empty.

Every connection Γ on $Y \rightarrow X$ is lifted to the connection

$$\gamma = \tilde{\Gamma} = dx^\alpha \otimes [\partial_\alpha + \Gamma_\alpha^i(y) \partial_i + (-\partial_j \Gamma_\alpha^i(y) p_i^\mu - K^\mu{}_{\nu\lambda}(x) p_j^\nu + K^\alpha{}_{\alpha\lambda}(x) p_j^\mu) \partial_\mu^j]$$

on $\Pi \rightarrow X$, where

$$K = dx^\alpha \otimes (\partial_\alpha + K^\mu{}_{\nu\lambda} \dot{x}_\mu \frac{\partial}{\partial \dot{x}_\nu})$$

is a linear symmetric connection on T^*X . We have the equality

$$\tilde{\Gamma} \rfloor \Omega = d(\Gamma \rfloor \theta).$$

This equality shows that $\tilde{\Gamma}$ is a Hamiltonian connection and

$$H_{\Gamma} = \Gamma \rfloor \theta = p_i^{\alpha} dy^i \wedge \omega_{\alpha} - p_i^{\alpha} \Gamma_{\alpha}^i \omega$$

is a Hamiltonian form.

Let H be a Hamiltonian form. For any exterior horizontal density $\tilde{H} = \tilde{\mathcal{H}}\omega$ on $\Pi \rightarrow X$, the form $H + \tilde{H}$ is a Hamiltonian form. Conversely, if H and H' are Hamiltonian forms, their difference $H - H'$ is an exterior horizontal density on $\Pi \rightarrow X$.

Thus, Hamiltonian forms constitute an affine space modelled on a linear space of the exterior horizontal densities on $\Pi \rightarrow X$. It follows that every Hamiltonian form on Π can be given by the expression (5.396) where Γ is some connection on $Y \rightarrow X$. Moreover, a Hamiltonian form has the canonical splitting (5.396) as follows.

Every Hamiltonian form H implies the momentum map

$$\hat{H} : \Pi \rightarrow J^1(X, Y), \quad y_{\alpha}^i \circ \hat{H} = \partial_{\alpha}^i \mathcal{H},$$

and the associated connection $\Gamma_H = \hat{H} \circ \hat{0}$ on Y where $\hat{0}$ is the global zero section of $\Pi \rightarrow Y$. As a consequence, we have the canonical splitting

$$H = H_{\Gamma_H} - \tilde{H}.$$

The Hamiltonian operator \mathcal{E}_H of a Hamiltonian form H is defined to be the first-order differential operator on $\Pi \rightarrow X$,

$$\begin{aligned} \mathcal{E}_H &: j^1\Pi \rightarrow \wedge^{n+1}T^*\Pi, \\ \mathcal{E}_H &= dH - \hat{\Omega} = [(y_{(\lambda)}^i - \partial_{\alpha}^i \mathcal{H}) dp_i^{\alpha} - (p_{i\lambda}^{\alpha} + \partial_i \mathcal{H}) dy^i] \wedge \end{aligned} \quad (5.413)$$

$$\text{where} \quad \hat{\Omega} = dp_i^{\alpha} \wedge dy^i \wedge \omega_{\alpha} + p_{i\lambda}^{\alpha} dy^i \wedge \omega - y_{(\lambda)}^i dp_i^{\alpha} \wedge \omega$$

is the pull-back of the multisymplectic form (5.394) onto $j^1\Pi$.

For any connection γ on $\Pi \rightarrow X$, we have

$$\mathcal{E}_H \circ \gamma = dH - \gamma \rfloor \Omega.$$

It follows that γ is a Hamiltonian connection for a Hamiltonian form H iff it takes its values into $\text{Ker } \mathcal{E}_H$ given by the coordinate relations

$$y_{(\lambda)}^i = \partial_{\alpha}^i \mathcal{H}, \quad p_{i\lambda}^{\alpha} = -\partial_i \mathcal{H}. \quad (5.414)$$

Let a Hamiltonian connection γ has an integral section r of $\Pi \rightarrow X$, that is, $\gamma \circ r = j^1r$. Then, the algebraic equations (5.414) are brought into the first-order differential Hamiltonian equations (5.397).

Now we consider relations between Lagrangian and Hamiltonian formalisms. A Hamiltonian form H is defined to be associated with a Lagrangian density L if it satisfies the relations

$$\begin{aligned}\widehat{L} \circ \widehat{H}|_Q &= Id_Q, & Q &= \widehat{L}(J^1(X, Y)), \\ H &= H_{\widehat{H}} + L \circ \widehat{H},\end{aligned}$$

which take the coordinate form

$$\partial_i^\mu \mathcal{L}(x^\alpha, y^j, \partial_\alpha^j \mathcal{H}) = p_i^\mu, \quad \mathcal{L}(x^\alpha, y^j, \partial_\alpha^j \mathcal{H}) = p_i^\mu \partial_\mu^i \mathcal{H} - \mathcal{H}.$$

Note that there are different Hamiltonian forms associated with the same singular Lagrangian density.

Bearing in mind physical application, we restrict our consideration to so-called *semiregular Lagrangian densities* L when the preimage $\widehat{L}^{-1}(q)$ of each point of $q \in Q$ is the connected submanifold of $J^1(X, Y)$. In this case, all Hamiltonian forms associated with a semiregular Lagrangian density L coincide on the Lagrangian constraint space Q , and the Poincaré–Cartan form Ξ_L is the pull-back

$$\Xi_L = H \circ \widehat{L}, \quad \pi_i^\alpha y_\alpha^i - \mathcal{L} = \mathcal{H}(x^\mu, y^i, \pi_i^\alpha),$$

of any associated multimomentum Hamiltonian form H by the Legendre morphism \widehat{L} [Zakharov (1992)]. Also the generating form (5.401) is the pull-back of

$$\Lambda_L = \mathcal{E}_H \circ J^1 \widehat{L}$$

of the Hamiltonian operator (5.413) of any Hamiltonian form H associated with a semiregular Lagrangian density L . As a consequence, we get the following correspondence between solutions of the Euler–Lagrangian equations and the Hamiltonian equations [Sardanashvily (1994); Zakharov (1992)].

Let a section r of $\Pi \rightarrow X$ be a solution of the Hamiltonian equations (5.397) for a Hamiltonian form H associated with a semiregular Lagrangian density L . If r lives on the Lagrangian constraint space Q , the section $\bar{s} = \widehat{H} \circ r$ of $J^1(X, Y) \rightarrow X$ satisfies the first-order Euler–Lagrangian equations (5.386). Conversely, given a semiregular Lagrangian density L , let \bar{s} be a solution of the first-order Euler–Lagrangian equations (5.386). Let H be a Hamiltonian form associated with L so that

$$\widehat{H} \circ \widehat{L} \circ \bar{s} = \bar{s}. \quad (5.415)$$

Then, the section $r = \widehat{L} \circ \bar{s}$ of $\Pi \rightarrow X$ is a solution of the Hamiltonian equations (5.397) for H . For sections \bar{s} and r , we have the relations

$$\bar{s} = j^1 s, \quad \text{and} \quad s = \pi_{\Pi Y} \circ r,$$

where s is a solution of the second-order Euler–Lagrangian equations (5.403).

We shall say that a family of Hamiltonian forms H associated with a semiregular Lagrangian density L is complete if, for each solution \bar{s} of the first-order Euler–Lagrangian equations (5.386), there exists a solution r of the Hamiltonian equations (5.397) for some Hamiltonian form H from this family so that

$$r = \widehat{L} \circ \bar{s}, \quad \bar{s} = \widehat{H} \circ r, \quad \bar{s} = J^1(\pi_{\Pi Y} \circ r). \quad (5.416)$$

Such a complete family exists iff, for each solution \bar{s} of the Euler–Lagrangian equations for L , there exists a Hamiltonian form H from this family so that the condition (5.415) holds.

We do not discuss here existence of solutions of Euler–Lagrangian and Hamiltonian equations. Note that, in contrast with mechanics, there are different Hamiltonian connections associated with the same multimomentum Hamiltonian form in general. Moreover, in field theory when the primary constraint space is the Lagrangian constraint space Q , there is a family of Hamiltonian forms associated with the same Lagrangian density as a rule. In practice, one can choose either the Hamiltonian equations or solutions of the Hamiltonian equations such that these solutions live on the constraint space.

Hamiltonian SEM–Tensors

Let H be a Hamiltonian form on the Legendre bundle Π over a fibre bundle $Y \rightarrow X$. We have the following differential conservation law on solutions of the Hamiltonian equations [Sardanashvily (1998)].

Let r be a section of the fibred Legendre manifold $\Pi \rightarrow X$. Given a connection Γ on $Y \rightarrow X$, we consider the T^*X –valued $(n-1)$ –form

$$\begin{aligned} T_\Gamma(r) &= -(\Gamma \rfloor H) \circ r, \\ T_\Gamma(r) &= [r_i^\alpha (\partial_\mu r^i - \Gamma_\mu^i) - \delta_\mu^\alpha (r_i^\alpha (\partial_\alpha r^i - \Gamma_\alpha^i) - \widetilde{\mathcal{H}}_\Gamma)] dx^\mu \otimes \omega_\alpha, \end{aligned} \quad (5.417)$$

on X where $\widetilde{\mathcal{H}}_\Gamma$ is the Hamiltonian density in the splitting (5.396) of H with respect to the connection Γ .

Let $\tau = \tau^\alpha \partial_\alpha$ be a vector-field on X . Given a connection Γ on $Y \rightarrow X$, it induces the projectable vector-field

$$\tilde{\tau}_\Gamma = \tau^\alpha \partial_\alpha + \tau^\alpha \Gamma_\alpha^i \partial_i + (-\tau^\mu p_j^\alpha \partial_i \Gamma_\mu^j - p_i^\alpha \partial_\mu \tau^\mu + p_i^\mu \partial_\mu \tau^\alpha) \partial_\alpha^i$$

on the Legendre bundle Π . Let us calculate the Lie derivative $\mathfrak{L}_{\tilde{\tau}_\Gamma} \tilde{H}_\Gamma$ on a section r . We have

$$\begin{aligned} (\mathfrak{L}_{\tilde{\tau}_\Gamma} \tilde{H}_\Gamma) \circ r &= \{ \partial_\alpha \tau^\alpha \tilde{\mathcal{H}}_\Gamma + [\tau^\alpha \partial_\alpha \\ &+ \tau^\alpha \Gamma_\alpha^i \partial_i + (-\tau^\mu r_j^\alpha \partial_i \Gamma_\mu^j - r_i^\alpha \partial_\mu \tau^\mu + r_i^\mu \partial_\mu \tau^\alpha) \partial_\alpha^i] \tilde{\mathcal{H}}_\Gamma \} \omega \\ &= \tau^\mu r_i^\alpha R_{\lambda\mu}^i \omega + d(\tau^\mu T_\Gamma^\alpha{}_\mu(r) \omega_\alpha) - (\tilde{\tau}_{\Gamma V} \rfloor \mathcal{E}_H) \circ r, \end{aligned} \quad (5.418)$$

where $\tilde{\tau}_{\Gamma V}$ is the vertical part of the canonical horizontal splitting (5.14) of the vector-field $\tilde{\tau}_V$ on Π over $j^1\Pi$. If r is a solution of the Hamiltonian equations, the equality (5.418) becomes the conservation law (5.398). The form (5.417) modulo the Hamiltonian equations reads

$$T_\Gamma(r) \approx [r_i^\alpha (\partial_\mu^i \mathcal{H} - \Gamma_\mu^i) - \delta_\mu^\alpha (r_i^\alpha \partial_\alpha^i \mathcal{H} - \mathcal{H})] dx^\mu \otimes \omega_\alpha. \quad (5.419)$$

For example, if $X = \mathbb{R}$ and Γ is the trivial connection, we have $T_\Gamma(r) = \mathcal{H} dt$, where \mathcal{H} is a Hamiltonian function. Then, the identity (5.398) becomes the conventional energy conservation law (5.384) in mechanics.

Unless $n = 1$, the identity (5.398) cannot be regarded directly as the energy-momentum conservation law. To clarify its physical meaning, we turn to the Lagrangian formalism.

Let a Hamiltonian form H be associated with a semiregular Lagrangian density L . Let r be a solution of the Hamiltonian equations of H which lives on the Lagrangian constraint space Q and \bar{s} the associated solution of the first-order Euler-Lagrangian equations of L so that they satisfy the conditions (5.416). Then, we have

$$T_\Gamma(r) = \mathcal{J}_\Gamma(\tilde{H} \circ r), \quad T_\Gamma(\tilde{L} \circ \bar{s}) = \mathcal{J}_\Gamma(\bar{s}),$$

where \mathcal{J}_Γ is the SEM-tensor (5.407).

It follows that, on the Lagrangian constraint space Q , the form (5.419) can be treated the Hamiltonian SEM-tensor relative to the connection Γ on $Y \rightarrow X$.

At the same time, the examples below show that, in several field models, the equality (5.398) is brought into the covariant conservation law (5.391) for the metric SEM-tensor.

In the Lagrangian formalism, the metric SEM-tensor is defined to be

$$\sqrt{-g}t_{\alpha\beta} = 2\frac{\partial\mathcal{L}}{\partial g^{\alpha\beta}}.$$

In case of a background world metric g , this object is well-behaved. In the framework of the multimomentum Hamiltonian formalism, one can introduce the similar tensor

$$\sqrt{-g}t_H^{\alpha\beta} = 2\frac{\partial\mathcal{H}}{\partial g_{\alpha\beta}}. \quad (5.420)$$

If a Hamiltonian form H is associated with a semiregular Lagrangian density L , there are the equalities

$$\begin{aligned} t_H^{\alpha\beta}(q) &= -g^{\alpha\mu}g^{\beta\nu}t_{\mu\nu}(x^\alpha, y^i, \partial_\alpha^i\mathcal{H}(q)), & (q \in Q), \\ t_H^{\alpha\beta}(x^\alpha, y^i, \pi_i^\alpha(z)) &= -g^{\alpha\mu}g^{\beta\nu}t_{\mu\nu}(z), & \widehat{H} \circ \widehat{L}(z) = z. \end{aligned}$$

In view of these equalities, we can think of the tensor (5.420) restricted to the Lagrangian constraint space Q as being the Hamiltonian metric SEM-tensor. On Q , the tensor (5.420) does not depend upon choice of a Hamiltonian form H associated with L . Therefore, we shall denote it by the common symbol t . Set

$$t^\lambda{}_\alpha = g_{\alpha\nu}t^{\lambda\nu}.$$

In the presence of a background world metric g , the identity (5.398) takes the form

$$t^\lambda{}_\alpha \{\alpha\lambda\mu\} \sqrt{-g} + (\Gamma_\mu^i \partial_i - \partial_i \Gamma_\mu^j r_j^\alpha \partial_\alpha^i) \widetilde{\mathcal{H}}_\Gamma \approx \frac{d}{dx^\alpha} T_\Gamma^\alpha{}_\mu + r_i^\alpha R_{\alpha\mu}^i, \quad (5.421)$$

where by $\{\alpha\lambda\mu\}$ are meant the Christoffel symbols of the world metric g .

SEM Tensors in Gauge Theory

In this subsection, following [Sardanashvily (1998)] we consider the gauge theory of principal connections treated as gauge potentials. Here, the manifold X is assumed to be oriented and provided with a nondegenerate fibre metric $g_{\mu\nu}$ in the tangent bundle of X . We denote $g = \det(g_{\mu\nu})$.

Let $P \rightarrow X$ be a principal bundle with a structure Lie group G which acts freely and transitively on P on the right: $r_g : p \mapsto pg$, $(p \in P, g \in G)$.

A principal connection A on $P \rightarrow X$ is defined to be a G -equivariant connection on P such that $j^1 r_g \circ A = A \circ r_g$ for each canonical morphism r_g .

Recall that there is the 1–1 correspondence between the principal connections on a principal bundle $P \rightarrow X$ and the global sections of the quotient bundle

$$C = J^1(X, P)/G \rightarrow X. \quad (5.422)$$

The bundle (5.422) is the affine bundle modelled on the vector bundle $\overline{C} = T^*X \otimes (V/P/G)$. Given a bundle atlas Ψ^P of P , the bundle C has the fibred coordinates (x^μ, k_μ^m) so that $(k_\mu^m \circ A)(x) = A_\mu^m(x)$ are coefficients of the local connection 1-form of a principal connection A with respect to the atlas Ψ^P . The 1-jet space $J^1(X, C)$ of the fibre bundle $C \rightarrow X$ is coordinated by $(x^\mu, k_\mu^m, k_{\mu\lambda}^m)$.

There exists the canonical splitting over C , given by

$$J^1(X, C) = C_+ \oplus C_- = (J^2P/G) \oplus (\wedge^2 T^*X \otimes V^G P), \quad (5.423)$$

$$k_{\mu\lambda}^m = \frac{1}{2}(k_{\mu\lambda}^m + k_{\lambda\mu}^m + c_{nl}^m k_\alpha^n k_\mu^l) + \frac{1}{2}(k_{\mu\lambda}^m - k_{\lambda\mu}^m - c_{nl}^m k_\alpha^n k_\mu^l).$$

The corresponding surjections read:

$$\begin{aligned} \mathcal{S} : J^1(X, C) &\rightarrow C_+, & \mathcal{S}_{\lambda\mu}^m &= k_{\mu\lambda}^m + k_{\lambda\mu}^m + c_{nl}^m k_\alpha^n k_\mu^l, \\ \mathcal{F} : J^1(X, C) &\rightarrow C_-, & \mathcal{F}_{\lambda\mu}^m &= k_{\mu\lambda}^m - k_{\lambda\mu}^m - c_{nl}^m k_\alpha^n k_\mu^l. \end{aligned}$$

On the configuration space (5.423), the conventional Yang–Mills Lagrangian density L_{YM} of gauge potentials in the presence of a background world metric is given by the expression

$$L_{YM} = \frac{1}{4\varepsilon^2} a_{mn}^G g^{\lambda\mu} g^{\beta\nu} \mathcal{F}_{\lambda\beta}^m \mathcal{F}_{\mu\nu}^n \sqrt{|g|} \omega, \quad (5.424)$$

where a^G is a nondegenerate G -invariant metric in the Lie algebra \mathfrak{g} of G . The Legendre morphism associated with the Lagrangian density (5.424) takes the form

$$p_m^{(\mu\lambda)} \circ \widehat{L}_{YM} = 0, \quad (5.425)$$

$$p_m^{[\mu\lambda]} \circ \widehat{L}_{YM} = \varepsilon^{-2} a_{mn}^G g^{\lambda\alpha} g^{\mu\beta} \mathcal{F}_{\alpha\beta}^n \sqrt{|g|}. \quad (5.426)$$

The equation (5.425) defines the constraint space of gauge theory.

Given a symmetric connection K on the tangent bundle TX , every principal connection B on P induces the connection

$$\Gamma_{\mu\lambda}^m = \partial_\mu B_\alpha^m - c_{nl}^m k_\mu^n B_\alpha^l - K^\beta{}_{\mu\lambda} (B_\beta^m - k_\beta^m) \quad (5.427)$$

on the bundle of principal connections C .

Let τ be a vector-field on the base X and

$$\tau_{BK} = \tau^\alpha \{ \partial_\alpha + [\partial_\mu B_\alpha^m - c_{nl}^m k_\mu^n B_\alpha^l - K^\beta_{\mu\lambda} (B_\beta^m - k_\beta^m)] \partial_m^\mu \} \quad (5.428)$$

its horizontal lift onto C by means of the connection (5.427). For every vector-field τ , one can choose the connection K on the tangent bundle TX which has τ as the geodesic field. In this case, the horizontal lift (5.428) of the vector-field τ becomes its canonical lift

$$\tau_B = \tau^\alpha \partial_\alpha + [\tau^\alpha (\partial_\mu B_\alpha^m - c_{nl}^m k_\mu^n B_\alpha^l) + \partial_\mu \tau^\alpha (B_\alpha^m - k_\alpha^m)] \partial_m^\mu, \quad (5.429)$$

by means of the principal connection B on the principal bundle P [Giacchetta and Mangiarotti (1990)]. The vector-field (5.429) is just the general principal vector-field on C that has been mentioned in the previous Section. Hence, the Lie derivative of the Lagrangian density (5.424) by the jet lift $\bar{\tau}_B$ of the field τ_B becomes

$$\mathcal{L}_{\bar{\tau}_B} L_{YM} = (\partial_\alpha \tau^\alpha \mathcal{L}_{YM} + \tau^\alpha \partial_\alpha \mathcal{L}_{YM} - \mathcal{F}_{\mu\nu}^m \partial_\alpha \tau^\mu \pi_m^{\nu\lambda}) \omega.$$

The corresponding SEM transformation law takes the form

$$\begin{aligned} \partial_\alpha \tau^\alpha \mathcal{L}_{YM} - \tau^\mu t_\beta^\alpha \sqrt{|g|} \Gamma_{\mu\alpha}^\beta - \mathcal{F}_{\mu\nu}^m \partial_\alpha \tau^\mu \pi_m^{\nu\lambda} \approx \\ \widehat{\partial}_\alpha [\pi_m^{\nu\lambda} (\tau^\mu (\partial_\nu B_\mu^m - c_{nl}^m k_\nu^n B_\mu^l) + \partial_\nu \tau^\mu (B_\mu^m - k_\mu^m) - \tau^\mu k_{\nu\mu}^m) + \delta_\mu^\alpha \tau^\mu \mathcal{L}_{YM}], \end{aligned} \quad (5.430)$$

where

$$t_\beta^\alpha = \frac{1}{\sqrt{|g|}} (\pi_m^{\nu\alpha} \mathcal{F}_{\beta\nu}^m - \delta_\beta^\alpha \mathcal{L}_{YM})$$

is the metric SEM-tensor of gauge potentials.

Note that, in general case of the principal connection B , the corresponding SEM transformation law (5.430) differs from the covariant conservation law in the Noether conservation law

$$\begin{aligned} \widehat{\partial}_\alpha (\pi_m^{\nu\lambda} u_{\mathfrak{g}\nu}^m) \approx 0, \quad \text{where} \\ u_{\mathfrak{g}} = (\partial_\nu \alpha^m + c_{nl}^m k_\nu^l \alpha^n) \partial_m^\nu, \quad \alpha^m = \tau^\mu (B_\mu^m - A_\mu^m) \end{aligned}$$

is the principal vector-field (5.353) on C .

Following the general procedure [Sardanashvily (1993); Sardanashvily (1994)], let us consider connections on the fibre bundle $C \rightarrow X$ which take their values into $\text{Ker } \hat{L}_{YM}$:

$$\Gamma : C \rightarrow C_+, \quad \Gamma_{\mu\lambda}^m - \Gamma_{\lambda\mu}^m - c_{nl}^m k_\alpha^n k_\mu^l = 0. \quad (5.431)$$

Moreover, we can restrict ourselves to connections of the following type. Every principal connection B on P induces the connection Γ_B (5.431) on C such that

$$\Gamma_B \circ B = \mathcal{S} \circ j^1 B,$$

$$\Gamma_{B\mu\lambda}^m = \frac{1}{2}[c_{nl}^m k_\alpha^n k_\mu^l + \partial_\mu B_\alpha^m + \partial_\alpha B_\mu^m - c_{nl}^m (k_\mu^n B_\alpha^l + k_\alpha^n B_\mu^l)] - \Gamma_{\mu\lambda}^\beta (B_{\beta'}^m - k_{\beta'}^m).$$

For all these connections, the following *Hamiltonian forms*

$$H_B = p_m^{\mu\lambda} dk_\mu^m \wedge \omega_\alpha - p_m^{\mu\lambda} \Gamma_{B\mu\lambda}^m \omega - \tilde{\mathcal{H}}_{YM} \omega, \quad (5.432)$$

$$\tilde{\mathcal{H}}_{YM} = \frac{\varepsilon^2}{4} a_G^{mn} g_{\mu\nu} g_{\lambda\beta} p_m^{[\mu\lambda]} p_n^{[\nu\beta]} |g|^{-1/2},$$

are associated with the Lagrangian density L_{YM} and constitute a complete family. The corresponding Hamiltonian equations for sections r of $\Pi \rightarrow X$ read

$$\partial_\alpha p_m^{\mu\lambda} = -c_{ln}^m k_\nu^l p_n^{[\mu\nu]} + c_{ml}^n B_\nu^l p_n^{(\mu\nu)} - \Gamma_{\lambda\nu}^\mu p_m^{(\lambda\nu)}, \quad (5.433)$$

$$\partial_\alpha k_\mu^m + \partial_\mu k_\alpha^m = 2\Gamma_{B(\mu\lambda)}^m, \quad (5.434)$$

plus the equation (5.426). The equations (5.426) and (5.433) restricted to the constraint space (5.425) are the familiar Yang–Mills equations. Different Hamiltonian forms (5.432) lead to the different equations (5.434). The equation (5.434) is independent of canonical momenta and plays the role of the gauge-type condition. Its solution is $k(x) = B$.

Let A be a solution of the Yang–Mills equations. There exists the Hamiltonian form $H_{B=A}$ (5.432) such that $r_A = \hat{L}_{YM} \circ A$ is a solution of the corresponding Hamiltonian equations (5.433), (5.434) and (5.426) on the constraint space (5.425).

On the solution r_A , the curvature of the connection Γ_A is reduced to

$$\begin{aligned} R_{\lambda\alpha\mu}^m &= \frac{1}{2}(\partial_\alpha F_{\alpha\mu}^m - c_{qn}^m k_\alpha^q F_{\alpha\mu}^n - \Gamma_{\alpha\lambda}^\beta F_{\beta\mu}^m - \Gamma_{\mu\lambda}^\beta F_{\alpha\beta}^m) = \\ &= \frac{1}{2}[(\partial_\alpha F_{\lambda\mu}^m - c_{qn}^m k_\alpha^q F_{\lambda\mu}^n - \Gamma_{\lambda\alpha}^\beta F_{\mu\beta}^m) - (\partial_\mu F_{\lambda\alpha}^m - c_{qn}^m k_\mu^q F_{\lambda\alpha}^n - \Gamma_{\lambda\mu}^\beta F_{\alpha\beta}^m)] \end{aligned}$$

where $F = \mathcal{F} \circ A$ is the strength of A . If we set

$$S^\alpha{}_\mu = p_m^{[\alpha\lambda]} \partial_{\alpha\mu}^m \tilde{\mathcal{H}}_{YM} = \frac{\varepsilon^2}{2\sqrt{|g|}} a_G^{mn} g_{\mu\nu} g_{\alpha\beta} p_m^{[\alpha\lambda]} p_n^{[\beta\nu]},$$

then we have

$$S^\alpha{}_\mu = \frac{1}{2} p^{[\alpha\lambda]} \mathcal{F}_{\mu\alpha}^m, \quad \tilde{\mathcal{H}}_{YM} = \frac{1}{2} S^\alpha{}_\alpha.$$

Using (5.425), (5.426) and (5.433), we get the relations

$$\partial_n^\beta \Gamma_{A\alpha\mu}^m p_m^{\alpha\lambda} \partial_{\beta\lambda}^n \tilde{\mathcal{H}}_{YM} = \Gamma_{\alpha\mu}^\beta S^\alpha{}_\beta, \quad r_{Am}^{[\lambda\alpha]} R_{\lambda\alpha\mu}^m = \partial_\alpha S^\alpha{}_\mu(r_A) - \Gamma_{\mu\lambda}^\beta S^\alpha{}_\beta(r_A)$$

and we find that

$$t^\alpha{}_\mu \sqrt{|g|} = 2S^\alpha{}_\mu - \frac{1}{2} \delta_\mu^\alpha S^\alpha{}_\alpha, \quad T_{\Gamma_A}{}^\alpha{}_\mu(r_A) = S^\alpha{}_\mu(r_A) - \frac{1}{2} \delta_\mu^\alpha S^\alpha{}_\alpha(r_A),$$

$$t^\alpha{}_\mu(r_A) \sqrt{|g|} = T_{\Gamma_A}{}^\alpha{}_\mu(r_A) + S^\alpha{}_\mu(r_A).$$

Hence, the identity (5.421) in gauge theory is brought into the covariant energy-momentum conservation law

$$\nabla_\alpha t^\alpha{}_\mu(r_A) \approx 0.$$

The Lagrangian partner of the Hamiltonian SEM-tensor $T_{\Gamma_A}(r_A)$ is the SEM-tensor $\mathcal{J}_{\Gamma_A}(A)$ (5.407) on the solution A relative to the connection Γ_A on the bundle C . This is exactly the familiar symmetrized canonical energy-momentum tensor of gauge potentials.

SEM Tensors of Matter Fields

In gauge theory, matter fields possessing only internal symmetries are described by sections of a vector bundle $Y = (P \times V)/G$, associated with a principal bundle P [Sardanashvily (1998)]. It has a G -invariant fibre metric a^Y . Because of the canonical vertical splitting $VY = Y \times Y$, the metric a^Y is a fibre metric in the vertical tangent bundle $VY \rightarrow Y$. Every principal connection A on a principal bundle P yields the associated connection

$$\Gamma = dx^\alpha \otimes [\partial_\alpha + A_\mu^m(x) I_m^i{}_j y^j \partial_i], \quad (5.435)$$

where $A_\mu^m(x)$ are coefficients of the local connection 1-form and I_m are generators of the structure group G on the standard fibre V of the bundle Y .

On the configuration space $J^1(X, Y)$, the regular Lagrangian density of matter fields in the presence of a background connection Γ on Y reads

$$L_{(m)} = \frac{1}{2} a_{ij}^Y [g^{\mu\nu} (y_\mu^i - \Gamma_\mu^i)(y_\nu^j - \Gamma_\nu^j) - m^2 y^i y^j] \sqrt{|g|} \omega. \quad (5.436)$$

The Legendre bundle of the vector bundle $Y \rightarrow X$ is $\Pi = \wedge^n T^*X \otimes TX \otimes Y^*$. The unique Hamiltonian form on Π associated with the La-

grangian density $L_{(m)}$ (5.436) is written

$$H_{(m)} = p_i^\alpha dy^i \wedge \omega_\alpha - p_i^\alpha \Gamma_\alpha^i \omega - \frac{1}{2} (a_Y^{ij} g_{\mu\nu} p_i^\mu p_j^\nu |g|^{-1} + m^2 a_{ij}^Y y^i y^j) \sqrt{|g|} \omega, \quad (5.437)$$

where a_Y is the fibre metric in V^*Y dual to a^Y . There is the 1–1 correspondence between the solutions of the first-order Euler–Lagrangian equations of the regular Lagrangian density (5.436) and the solutions of the Hamiltonian equations of the Hamiltonian form (5.437).

To examine the conservation law (5.421), let us take the same Hamiltonian SEM–tensor relative to the connection Γ (5.435) for all solutions r of the Hamiltonian equations. The following equality motivates the option above. We have

$$\begin{aligned} T_{\Gamma\mu}^\alpha(r) &= t^\alpha{}_\mu(r) \sqrt{|g|} = [a_Y^{ij} g_{\mu\nu} r_i^\alpha p_j^\nu |g|^{-1} \\ &\quad - \frac{1}{2} \delta_\mu^\alpha (a_Y^{ij} g_{\alpha\nu} r_i^\alpha r_j^\nu |g|^{-1} + m^2 a_{ij}^Y r^i r^j)] \sqrt{|g|}. \end{aligned}$$

The gauge invariance condition $I_m^j r_j^\alpha \partial_\alpha^i \tilde{\mathcal{H}} = 0$ also takes place. Then, it can be observed that the identity (5.421) reduces to the familiar covariant energy–momentum conservation law

$$\sqrt{|g|} \nabla_\alpha t^\alpha{}_\mu(r) \approx -r_i^\alpha F_{\lambda\mu}^m I_m^i y^j.$$

SEM Tensors in Affine–Metric Gravitation Theory

Now we can apply the Hamiltonian SEM–tensor machinery to gravitation theory [Sardanashvily (1998); Giachetta and Sardanashvily (1996)]. Here, X^4 is a 4D world manifold which obeys the well-known topological conditions in order that a gravitational field exists on X^4 .

Recall that the contemporary concept of gravitational interaction is based on the gauge gravitation theory with two types of gravitational fields: tetrad gravitational fields and Lorentz gauge potentials. In absence of fermion matter, one can choose by gravitational variables a pseudo–Riemannian metric g on a world space–time manifold X^4 and a general linear connections K on the tangent bundle of X^4 . We call them a world metric and a world connection respectively. Here we are not concerned with the matter interacting with a general linear connection, for it is non–Lagrangian and hypothetical as a rule.

Let $LX \rightarrow X^4$ be the principal bundle of linear frames in the tangent spaces to X^4 . Its structure group is $GL^+(4, \mathbb{R})$. The world connections are

associated with the principal connections on the principal bundle $LX \rightarrow X^4$. Hence, there is the 1–1 correspondence between the world connections and the global sections of the quotient bundle

$$C = J^1(X^4, LX)/GL^+(4, \mathbb{R}). \quad (5.438)$$

We therefore can apply the standard procedure of the Hamiltonian gauge theory in order to describe the configuration and phase-spaces of world connections [Sardanashvily (1993); Sardanashvily (1994)].

Also, there is the 1–1 correspondence between the world metrics g on X^4 and the global sections of the bundle Σ of pseudo-Riemannian bilinear forms in tangent spaces to X^4 . This bundle is associated with the GL_4 –principal bundle LX . The 2-fold covering of the bundle Σ is the quotient bundle $LX/SO(3, 1)$.

The total configuration space of the affine-metric gravitational variables is the product

$$J^1(X^4, C) \times J^1(X^4, \Sigma). \quad (5.439)$$

coordinated by $(x^\mu, g^{\alpha\beta}, k^\alpha_{\beta\mu}, g^{\alpha\beta}_{\alpha}, k^\alpha_{\beta\mu\lambda})$. Also, the total phase-space Π of the affine-metric gravity is the product of the Legendre bundles over the above-mentioned bundles C and Σ . It has the corresponding canonical coordinates $(x^\mu, g^{\alpha\beta}, k^\alpha_{\beta\mu}, p_{\alpha\beta}{}^\alpha, p_{\alpha}{}^{\beta\mu\lambda})$.

On the configuration space (5.439), the Hilbert–Einstein Lagrangian density of general relativity reads

$$L_{HE} = -\frac{1}{2\kappa} g^{\beta\lambda} \mathcal{F}^\alpha_{\beta\alpha\lambda} \sqrt{-g} \omega, \quad \text{with} \quad (5.440)$$

$$\mathcal{F}^\alpha_{\beta\nu\lambda} = k^\alpha_{\beta\lambda\nu} - k^\alpha_{\beta\nu\lambda} + k^\alpha_{\varepsilon\nu} k^\varepsilon_{\beta\lambda} - k^\alpha_{\varepsilon\lambda} k^\varepsilon_{\beta\nu}.$$

The corresponding Legendre morphism is given by the expressions

$$p_{\alpha\beta}{}^\alpha \circ \widehat{L}_{HE} = 0,$$

$$p_{\alpha}{}^{\beta\nu\lambda} \circ \widehat{L}_{HE} = \pi_{\alpha}{}^{\beta\nu\lambda} = \frac{1}{2\kappa} (\delta^\nu_{\alpha} g^{\beta\lambda} - \delta^\alpha_{\alpha} g^{\beta\nu}) \sqrt{-g}, \quad (5.441)$$

which define the constraint space of general relativity in the affine-metric variables.

Now, let us consider the following connections on the bundle $C \times \Sigma$ in order to construct a complete family of Hamiltonian forms associated with the Lagrangian density (5.440).

Let K be a world space-time connection and

$$\Gamma_{K^\alpha \beta \nu \lambda} = \frac{1}{2} [k^\alpha_{\varepsilon \nu} k^\varepsilon_{\beta \lambda} - k^\alpha_{\varepsilon \lambda} k^\varepsilon_{\beta \nu} + \partial_\alpha K^\alpha_{\beta \nu} + \partial_\nu K^\alpha_{\beta \lambda} - 2K^\varepsilon_{(\nu \lambda)} (K^\alpha_{\beta \varepsilon} - k^\alpha_{\beta \varepsilon}) + K^\varepsilon_{\beta \lambda} k^\alpha_{\varepsilon \nu} + K^\varepsilon_{\beta \nu} k^\alpha_{\varepsilon \lambda} - K^\alpha_{\varepsilon \lambda} k^\varepsilon_{\beta \nu} - K^\alpha_{\varepsilon \nu} k^\varepsilon_{\beta \lambda}]$$

be the corresponding connection on the bundle C (5.438). Let K' be another symmetric world connection. Building on these connections, we set up the following connection on the bundle $C \times \Sigma$,

$$\begin{aligned} \Gamma^{\alpha\beta}{}_\alpha &= -K'^\alpha{}_{\varepsilon\lambda} g^{\varepsilon\beta} - K'^\beta{}_{\varepsilon\lambda} g^{\alpha\varepsilon}, \\ \Gamma^\alpha{}_{\beta\nu\lambda} &= \Gamma_{K^\alpha \beta \nu \lambda} - \frac{1}{2} R^\alpha{}_{\beta\nu\lambda}, \end{aligned} \quad (5.442)$$

where $R^\alpha{}_{\beta\nu\lambda}$ is the *Riemann curvature tensor* of K .

For all connections (5.442), the following Hamiltonian forms are associated with the Lagrangian density L_{HE} and constitute a complete family:

$$\begin{aligned} H_{HE} &= (p_{\alpha\beta}{}^\alpha dg^{\alpha\beta} + p_\alpha{}^{\beta\nu\lambda} dk^\alpha_{\beta\nu}) \wedge \omega_\alpha - \mathcal{H}_{HE} \omega, \\ \mathcal{H}_{HE} &= -p_{\alpha\beta}{}^\alpha (K'^\alpha{}_{\varepsilon\lambda} g^{\varepsilon\beta} + K'^\beta{}_{\varepsilon\lambda} g^{\alpha\varepsilon}) \\ &\quad + p_\alpha{}^{\beta\nu\lambda} \Gamma_{K^\alpha \beta \nu \lambda} - \frac{1}{2} R^\alpha{}_{\beta\nu\lambda} (p_\alpha{}^{\beta\nu\lambda} - \pi_\alpha{}^{\beta\nu\lambda}) \\ &= -p_{\alpha\beta}{}^\alpha (K'^\alpha{}_{\varepsilon\lambda} g^{\varepsilon\beta} + K'^\beta{}_{\varepsilon\lambda} g^{\alpha\varepsilon}) + p_\alpha{}^{\beta\nu\lambda} \Gamma^\alpha{}_{\beta\nu\lambda} + \tilde{\mathcal{H}}_{HE}, \\ \tilde{\mathcal{H}}_{HE} &= \frac{1}{2\kappa} R \sqrt{-g}. \end{aligned} \quad (5.443)$$

Given the Hamiltonian form H_{HE} (5.443) plus a Hamiltonian form H_M for matter, we have the corresponding Hamiltonian equations

$$\partial_\alpha g^{\alpha\beta} + K'^\alpha{}_{\varepsilon\lambda} g^{\varepsilon\beta} + K'^\beta{}_{\varepsilon\lambda} g^{\alpha\varepsilon} = 0, \quad (5.444)$$

$$\partial_\alpha k^\alpha_{\beta\nu} = \Gamma_{K^\alpha \beta \nu \lambda} - \frac{1}{2} R^\alpha{}_{\beta\nu\lambda}, \quad (5.445)$$

$$\partial_\alpha p_{\alpha\beta}{}^\alpha = p_{\varepsilon\beta}{}^\sigma K'^\varepsilon{}_{\alpha\sigma} + p_{\varepsilon\alpha}{}^\sigma K'^\varepsilon{}_{\beta\sigma} \quad (5.446)$$

$$\begin{aligned} -\frac{1}{2\kappa} (R_{\alpha\beta} - \frac{1}{2} g_{\alpha\beta} R) \sqrt{-g} - \frac{\partial \mathcal{H}_M}{\partial g^{\alpha\beta}}, \\ \partial_\alpha p_\alpha{}^{\beta\nu\lambda} = -p_\alpha{}^{\varepsilon[\nu\gamma]} k^\beta_{\varepsilon\gamma} + p_\varepsilon{}^{\beta[\nu\gamma]} k^\varepsilon_{\alpha\gamma} - p_\alpha{}^{\beta\varepsilon\gamma} K^\nu_{(\varepsilon\gamma)} \\ - p_\alpha{}^{\varepsilon(\nu\gamma)} K^\beta_{\varepsilon\gamma} + p_\varepsilon{}^{\beta(\nu\gamma)} K^\varepsilon_{\alpha\gamma}, \end{aligned} \quad (5.447)$$

plus the motion equations of matter. The Hamiltonian equations (5.444) and (5.445) are independent of canonical momenta and so, reduce to the gauge-type conditions. The equation (5.445) breaks into the following two

parts,

$$F^\alpha{}_{\beta\lambda\nu} = R^\alpha{}_{\beta\nu\lambda}, \quad \text{and} \quad (5.448)$$

$$\begin{aligned} \partial_\nu(K^\alpha{}_{\beta\lambda} - k^\alpha{}_{\beta\lambda}) + \partial_\alpha(K^\alpha{}_{\beta\nu} - k^\alpha{}_{\beta\nu}) - 2K^\varepsilon{}_{(\nu\lambda)}(K^\alpha{}_{\beta\varepsilon} - k^\alpha{}_{\beta\varepsilon}) \\ + K^\varepsilon{}_{\beta\lambda}k^\alpha{}_{\varepsilon\nu} + K^\varepsilon{}_{\beta\nu}k^\alpha{}_{\varepsilon\lambda} - K^\alpha{}_{\varepsilon\lambda}k^\varepsilon{}_{\beta\nu} - K^\alpha{}_{\varepsilon\nu}k^\varepsilon{}_{\beta\lambda} = 0, \end{aligned} \quad (5.449)$$

where F is the curvature of the connection $k(x)$. It is clear that the gauge-type conditions (5.444) and (5.445) are satisfied by

$$k(x) = K, \quad K'^\alpha{}_{\beta\lambda} = \Gamma^\alpha{}_{\beta\lambda}. \quad (5.450)$$

When restricted to the constraint space (5.441), the Hamiltonian equations (5.446) and (5.447) become

$$\frac{1}{\kappa}(R_{\alpha\beta} - \frac{1}{2}g_{\alpha\beta}R)\sqrt{-g} = -\frac{\partial\mathcal{H}_M}{\partial g^{\alpha\beta}}, \quad (5.451)$$

$$\begin{aligned} D_\alpha(\sqrt{-g}g^{\nu\beta}) - \delta^\nu_\alpha D_\alpha(\sqrt{-g}g^{\lambda\beta}) + \sqrt{-g}[g^{\nu\beta}(k^\alpha{}_{\alpha\lambda} - k^\alpha{}_{\lambda\alpha}) \\ + g^{\lambda\beta}(k^\nu{}_{\lambda\alpha} - k^\nu{}_{\alpha\lambda}) + \delta^\nu_\alpha g^{\lambda\beta}(k^\mu{}_{\mu\lambda} - k^\mu{}_{\lambda\mu})] = 0, \end{aligned} \quad (5.452)$$

$$\text{where} \quad D_\alpha g^{\alpha\beta} = \partial_\alpha g^{\alpha\beta} + k^\alpha{}_{\mu\lambda}g^{\mu\beta} + k^\beta{}_{\mu\lambda}g^{\alpha\mu}.$$

Substituting the equation (5.448) into the equation (5.451), we get the *Einstein equations*

$$\frac{1}{\kappa}(F_{\alpha\beta} - \frac{1}{2}g_{\alpha\beta}F) = -t_{\alpha\beta}, \quad (5.453)$$

where $t_{\alpha\beta}$ is the metric SEM-tensor of matter. The equations (5.452) and (5.453) are the familiar equations of affine-metric gravity. In particular, the former is the equation for torsion and nonmetricity terms of the general linear connection $k(x)$. In the absence of matter sources of a general linear connection, it admits the well-known solution

$$k^\alpha{}_{\beta\nu} = \Gamma^\alpha{}_{\beta\nu} - \frac{1}{2}\delta^\alpha_\nu V_\beta, \quad D_\alpha g^{\beta\gamma} = V_\alpha g^{\beta\gamma},$$

where V_α is an arbitrary covector-field corresponding to the well-known projective freedom.

Let $s = (k(x), g(x))$ be a solution of the Euler-Lagrangian equations of the first-order Hilbert-Einstein Lagrangian density (5.440) and r the corresponding solution of the Hamiltonian equations of the Hamiltonian form (5.443) where K and K' are given by the expressions (5.450). For this solution r , let us take the SEM-tensor T_s (5.399) relative to the connection

(5.442) where K and K' are given by the expressions (5.450). It reads

$$T_s^\alpha{}_\mu = \delta_\mu^\alpha \tilde{\mathcal{H}}_{HE} = \frac{1}{2\kappa} \delta_\mu^\alpha R \sqrt{-g}$$

and the identity (5.398) takes the form

$$\begin{aligned} & (\partial_\mu + \Gamma^{\alpha\beta}{}_\mu \partial_{\alpha\beta} + \Gamma_\mu^i \partial_i - \partial_i \Gamma_\mu^j p_j^\alpha \partial_\lambda^i)(\tilde{\mathcal{H}}_{HE} + \tilde{\mathcal{H}}_M) \\ & \approx \frac{d}{dx^\alpha} (T_s^\alpha{}_\mu + T_M^\alpha{}_\mu) + p_\alpha^{\beta\nu\lambda} R^\alpha{}_{\beta\nu\lambda\mu} + p_i^\alpha R^\lambda_{\lambda\mu} \end{aligned} \quad (5.454)$$

where T_M is the SEM-tensor for matter.

One can verify that the SEM-tensor T_s meets the condition

$$(\partial_\mu + \Gamma^{\alpha\beta}{}_\mu \partial_{\alpha\beta}) \tilde{\mathcal{H}}_{HE} = \frac{d}{dx^\alpha} T_s^\alpha{}_\mu, \quad (5.455)$$

so that on solutions (5.450), the curvature of the connection (5.442) vanishes. Hence, the identity (5.454) is reduced to the conservation law (5.421) of matter in the presence of a background metric. The gravitation SEM-tensor is eliminated from the conservation law because the Hamiltonian form \mathcal{H}_{HE} is affine in all canonical momenta. Note that only gauge-type conditions (5.444), (5.445) and the motion equations of matter have been used.

At the same time, since the canonical momenta $p_{\alpha\beta}{}^\alpha$ of the world metric are equal to zero, the Hamiltonian equation (5.446) on the Lagrangian constraint space becomes

$$\partial_{\alpha\beta}(\tilde{\mathcal{H}}_{HE} + \tilde{\mathcal{H}}_M) = 0.$$

Hence, the equality (5.454) takes the form

$$\pi_\alpha^{\beta\nu\lambda} \partial_\mu R^\alpha{}_{\beta\nu\lambda} + (\partial_\mu + \Gamma_\mu^i \partial_i - \partial_i \Gamma_\mu^j p_j^\alpha \partial_\lambda^i) \tilde{\mathcal{H}}_M \approx \frac{d}{dx^\alpha} (T_s^\alpha{}_\mu + T_M^\lambda{}_\mu) + p_i^\alpha R^\lambda_{\lambda\mu}. \quad (5.456)$$

This is the form of the energy-momentum conservation law which we observe also in case of quadratic Lagrangian densities of affine-metric gravity. Substituting the equality (5.455) into (5.456), we get the above result.

As a test case of quadratic Lagrangian densities of affine-metric gravity, let us examine the sum

$$L = \left(-\frac{1}{2\kappa} g^{\beta\lambda} \mathcal{F}^\alpha{}_{\beta\alpha\lambda} + \frac{1}{4\varepsilon} g_{\alpha\gamma} g^{\beta\sigma} g^{\nu\mu} g^{\lambda\varepsilon} \mathcal{F}^\alpha{}_{\beta\nu\lambda} \mathcal{F}^\gamma{}_{\sigma\mu\varepsilon} \right) \sqrt{-g} \omega \quad (5.457)$$

of the Hilbert–Einstein Lagrangian density and the Yang–Mills one. The corresponding Legendre map reads

$$p_{\alpha\beta}{}^{\alpha} \circ \widehat{L} = 0, \quad (5.458)$$

$$p_{\alpha}{}^{\beta(\nu\lambda)} \circ \widehat{L} = 0, \quad (5.459)$$

$$p_{\alpha}{}^{\beta[\nu\lambda]} \circ \widehat{L} = \pi_{\alpha}{}^{\beta\nu\lambda} + \frac{1}{\varepsilon} g_{\alpha\gamma} g^{\beta\sigma} g^{\nu\mu} g^{\lambda\varepsilon} \mathcal{F}^{\gamma}{}_{\sigma\varepsilon\mu} \sqrt{-g}. \quad (5.460)$$

The relations (5.458) and (5.459) defines the Lagrangian constraint space.

Let us consider two connections on the bundle $C \times \Sigma$,

$$\Gamma^{\alpha\beta}{}_{\alpha} = -K'^{\alpha}{}_{\varepsilon\lambda} g^{\varepsilon\beta} - K'^{\beta}{}_{\varepsilon\lambda} g^{\alpha\varepsilon}, \quad \text{and} \quad \Gamma^{\alpha}{}_{\beta\nu\lambda} = \Gamma_{K^{\alpha}}{}_{\beta\nu\lambda}, \quad (5.461)$$

where the notations of the expression (5.442) are used. The corresponding Hamiltonian forms

$$\begin{aligned} H &= (p_{\alpha\beta}{}^{\alpha} dg^{\alpha\beta} + p_{\alpha}{}^{\beta\nu\lambda} dk^{\alpha}{}_{\beta\nu}) \wedge \omega_{\alpha} - \mathcal{H}\omega, \\ \mathcal{H} &= -p_{\alpha\beta}{}^{\alpha} (K'^{\alpha}{}_{\varepsilon\lambda} g^{\varepsilon\beta} + K'^{\beta}{}_{\varepsilon\lambda} g^{\alpha\varepsilon}) + p_{\alpha}{}^{\beta\nu\lambda} \Gamma_{K^{\alpha}}{}_{\beta\nu\lambda} + \widetilde{\mathcal{H}}, \\ \widetilde{\mathcal{H}} &= \frac{\varepsilon}{4} g^{\alpha\gamma} g_{\beta\sigma} g_{\nu\mu} g_{\lambda\varepsilon} (p_{\alpha}{}^{\beta[\nu\lambda]} - \pi_{\alpha}{}^{\beta\nu\lambda}) (p_{\gamma}{}^{\sigma[\mu\varepsilon]} - \pi_{\gamma}{}^{\beta\mu\varepsilon}), \end{aligned} \quad (5.462)$$

are associated with the Lagrangian density (5.457) and constitute a complete family.

Given the Hamiltonian form (5.462) plus the Hamiltonian form H_M for matter, we have the corresponding Hamiltonian equations

$$\partial_{\alpha} g^{\alpha\beta} + K'^{\alpha}{}_{\varepsilon\lambda} g^{\varepsilon\beta} + K'^{\beta}{}_{\varepsilon\lambda} g^{\alpha\varepsilon} = 0, \quad (5.463)$$

$$\partial_{\alpha} k^{\alpha}{}_{\beta\nu} = \Gamma_{K^{\alpha}}{}_{\beta\nu\lambda} + \varepsilon g^{\alpha\gamma} g_{\beta\sigma} g_{\nu\mu} g_{\lambda\varepsilon} (p_{\gamma}{}^{\sigma[\mu\varepsilon]} - \pi_{\gamma}{}^{\beta\mu\varepsilon}), \quad (5.464)$$

$$\partial_{\alpha} p_{\alpha\beta}{}^{\alpha} = -\frac{\partial \mathcal{H}}{\partial g^{\alpha\beta}} - \frac{\partial \mathcal{H}_M}{\partial g^{\alpha\beta}}, \quad (5.465)$$

$$\begin{aligned} \partial_{\alpha} p_{\alpha}{}^{\beta\nu\lambda} &= -p_{\alpha}{}^{\varepsilon[\nu\gamma]} k^{\beta}{}_{\varepsilon\gamma} + p_{\varepsilon}{}^{\beta[\nu\gamma]} k^{\varepsilon}{}_{\alpha\gamma} \\ &\quad - p_{\alpha}{}^{\beta\varepsilon\gamma} K^{\nu}{}_{(\varepsilon\gamma)} - p_{\alpha}{}^{\varepsilon(\nu\gamma)} K^{\beta}{}_{\varepsilon\gamma} + p_{\varepsilon}{}^{\beta(\nu\gamma)} K^{\varepsilon}{}_{\alpha\gamma} \end{aligned} \quad (5.466)$$

plus the motion equations for matter. The equation (5.464) breaks into the equation (5.460) and the gauge-type condition (5.449). The gauge-type conditions (5.463) and (5.449) have the solution (5.450). Substituting the equation (5.464) into the equation (5.465) on the constraint space (5.458), we get the quadratic Einstein equations. Substitution of the equations (5.459) and (5.460) into the equation (5.466) results into the Yang–Mills generalization of the equation (5.452),

$$\partial_{\alpha} p_{\alpha}{}^{\beta\nu\lambda} + p_{\alpha}{}^{\varepsilon[\nu\gamma]} k^{\beta}{}_{\varepsilon\gamma} - p_{\varepsilon}{}^{\beta[\nu\gamma]} k^{\varepsilon}{}_{\alpha\gamma} = 0.$$

Consider now the splitting of the Hamiltonian form (5.462) with respect to the connection (5.442) and the Hamiltonian density

$$\tilde{\mathcal{H}}_\Gamma = \tilde{\mathcal{H}} + \frac{1}{2} p_\alpha^{\beta\nu\lambda} R^\alpha_{\beta\nu\lambda}.$$

Let $s = (k(x), g(x))$ be a solution of the Euler–Lagrangian equations of the Lagrangian density (5.457) and r the corresponding solution of the Hamiltonian equations of the Hamiltonian form (5.462) where K and K' are given by the expressions (5.450). For this solution r , let us take the SEM–tensor T_s (5.399) relative to the connection (5.442) where K and K' are given by the expressions (5.450). It reads

$$\begin{aligned} T_s^\alpha{}_\mu &= \frac{1}{2} p_\alpha^{\beta[\nu\lambda]} R^\alpha_{\beta\nu\mu} + \frac{\varepsilon}{2} g^{\alpha\gamma} g_{\beta\sigma} g_{\nu\delta} g_{\mu\varepsilon} p_\alpha^{\beta[\nu\lambda]} (p_\gamma^{\sigma[\delta\varepsilon]} - \pi_\gamma^{\sigma\delta\varepsilon}) \\ &\quad - \delta_\mu^\alpha (\tilde{\mathcal{H}} + \frac{\varepsilon}{2} g^{\alpha\gamma} g_{\beta\sigma} g_{\nu\delta} g_{\tau\varepsilon} \pi_\alpha^{\beta\nu\tau} (p_\gamma^{\sigma[\delta\varepsilon]} - \pi_\gamma^{\sigma\delta\varepsilon})) \end{aligned}$$

and is equal to

$$\frac{1}{\varepsilon} R_\alpha^{\beta\nu\lambda} R^\alpha_{\beta\nu\mu} + \pi_\alpha^{\beta\nu\lambda} R^\alpha_{\beta\nu\mu} - \delta_\mu^\alpha \left(\frac{1}{4\varepsilon} R_\alpha^{\beta\nu\lambda} R^\alpha_{\beta\nu\lambda} + \frac{1}{2\kappa} R \right).$$

The weak identity (5.398) now becomes

$$\begin{aligned} (\partial_\mu + \Gamma_\mu^{\alpha\beta} \partial_{\alpha\beta} + \Gamma_\mu^i \partial_i - \partial_i \Gamma_\mu^j p_j^\alpha \partial_\lambda^i - p_\alpha^{\beta\nu\lambda} \frac{\partial}{\partial k^\sigma{}_{\gamma\delta}} \Gamma_{K^\alpha}{}_{\beta\nu\mu} \frac{\partial}{\partial P_\sigma{}_{\gamma\delta\lambda}}) (\tilde{\mathcal{H}}_\Gamma + \tilde{\mathcal{H}}_M) \\ \approx \frac{d}{dx^\alpha} (T_s^\alpha{}_\mu + T_M^\alpha{}_\mu) + p_\alpha^{\beta\nu\lambda} R^\alpha_{\beta\nu\lambda\mu} + p_i^\alpha R_{\lambda\mu}^i \end{aligned}$$

and can be simplified to

$$\begin{aligned} (\partial_\mu + \Gamma_\mu^i \partial_i - \partial_i \Gamma_\mu^j p_j^\alpha \partial_\lambda^i) \tilde{\mathcal{H}}_M - p_\alpha^{\beta\nu\lambda} \frac{\partial}{\partial k^\sigma{}_{\gamma\delta}} \Gamma_{K^\alpha}{}_{\beta\nu\mu} \frac{\partial}{\partial p_\sigma{}_{\gamma\delta\lambda}} \tilde{\mathcal{H}}_\Gamma \\ \approx \frac{d}{dx^\alpha} (T_s^\alpha{}_\mu + T_M^\alpha{}_\mu) + p_i^\alpha R_{\lambda\mu}^i, \quad \text{where} \end{aligned} \quad (5.467)$$

$$\begin{aligned} p_\alpha^{\beta\nu\lambda} \frac{\partial}{\partial k^\sigma{}_{\gamma\delta}} \Gamma_{K^\alpha}{}_{\beta\nu\mu} \frac{\partial}{\partial p_\sigma{}_{\gamma\delta\lambda}} \tilde{\mathcal{H}}_\Gamma \\ = \frac{1}{\kappa} k^\gamma{}_{\beta\mu} (g^{\beta\nu} R^\alpha_{\gamma\alpha\nu} - g^{\alpha\nu} R^\beta_{\alpha\nu\gamma}) \sqrt{-g} - k^\gamma{}_{(\beta\mu)} p_\alpha^{\nu\beta\lambda} R^\alpha_{\nu\gamma\lambda}. \end{aligned} \quad (5.468)$$

Let us choose the local geodetic coordinate system at a point $x \in X$. Relative to this coordinate system, the equality (5.467) at x becomes the conservation law

$$(\partial_\mu + \Gamma_\mu^i \partial_i - \partial_i \Gamma_\mu^j p_j^\alpha \partial_\lambda^i) \tilde{\mathcal{H}}_M \approx \frac{d}{dx^\alpha} (T_s^\alpha{}_\mu + T_M^\alpha{}_\mu) + p_i^\alpha R_{\lambda\mu}^i.$$

For example, in gauge theory, we have

$$\frac{d}{dx^\alpha}(T_\Gamma^\alpha{}_\mu + t_M^\alpha{}_\mu) = 0,$$

where t_M is the metric SEM-tensor of matter.

5.12.2 Gauge Systems of Gravity and Fermion Fields

In physical reality, one observes three types of field systems: gravitational fields, fermion fields, and gauge fields associated with internal symmetries (see [Giachetta and Sardanashvily (1997)]). If the gauge invariance under internal symmetries is kept in the presence of a gravitational field, Lagrangian densities of gauge fields must depend on a metric gravitational field only.

In the gauge gravitation theory, gravity is represented by pairs (h, A_h) of gravitational fields h and associated Lorentz connections A_h [Hehl *et. al.* (1995); Sardanashvily (1992)]. The connection A_h is usually identified with both a connection on a world manifold X and a *spinor connection* on the *spinor bundle* $S_h \rightarrow X$ whose sections describe Dirac fermion fields ψ_h in the presence of the gravitational field h . The problem arises when Dirac fermion fields are described in the framework of the affine-metric gravitation theory. In this case, the fact that a world connection is some Lorentz connection may result from the field equations, but it cannot be assumed in advance. There are models where the world connection is not a Lorentz connection [Hehl *et. al.* (1995)]. Moreover, it may happen that a world connection is the Lorentz connection with respect to different gravitational fields [Thompson (1993)]. At the same time, a Dirac fermion field can be regarded only in a pair (h, ψ_h) with a certain gravitational field h .

One has to define the representation of cotangent vectors to X by the Dirac's γ -matrices in order to construct the Dirac operator. Given a *tetrad gravitational field* $h(x)$, we have the representation

$$\gamma_h : dx^\mu \mapsto \widehat{dx}^\mu = h_a^\mu \gamma^a.$$

However, different gravitational fields h and h' yield the nonequivalent representations γ_h and $\gamma_{h'}$.

It follows that fermion-gravitation pairs (h, ψ_h) are described by sections of the *composite spinor bundle*

$$S \rightarrow \Sigma \rightarrow X, \quad (5.469)$$

where $\Sigma \rightarrow X$ is the bundle of gravitational fields h where values of h play the role of parameter coordinates, besides the familiar world coordinates [Sardanashvily (1992)]. In particular, every spinor bundle $S_h \rightarrow X$ is isomorphic to the restriction of $S \rightarrow \Sigma$ to $h(X) \subset \Sigma$. Performing this restriction, we come to the familiar case of a field model in the presence of a gravitational field $h(x)$. The feature of the dynamics of field systems on the composite bundle (5.469) lies in the fact that we have the modified covariant differential of fermion fields which depend on derivatives of gravitational fields h .

As a consequence, we get the following *covariant derivative of Dirac fermion fields* in the presence of a gravitational field $h(x)$:

$$\begin{aligned}\tilde{D}_\alpha &= \partial_\alpha - \frac{1}{2} A^{abc}_\mu (\partial_\alpha h_c^\mu + K^\mu_{\nu\lambda} h_c^\nu) I_{ab}, \\ A^{abc}_\mu &= \frac{1}{2} (\eta^{ca} h_\mu^b - \eta^{cb} h_\mu^a),\end{aligned}\tag{5.470}$$

where K is a general linear connection on a world manifold X ,⁵ η is the Minkowski metric, and $I_{ab} = \frac{1}{4}[\gamma_a, \gamma_b]$ are generators of the *spinor Lie group* $L_s = SL(2, \mathbb{C})$.

The covariant derivative (5.470) has been considered by [Aringazin and Mikhailov (1991); Ponomarev and Obukhov (1982); Tucker and Wang (1995)]. The relation (5.472) correspond to the canonical decomposition of the Lie algebra of the general linear group. By the well-known Theorem [Kobayashi and Nomizu (1963/9)], every general linear connection being projected onto the Lie algebra of the Lorentz group induces a Lorentz connection.

In our opinion, the advantage of the covariant derivative (5.470), consists in the fact that, being derived in the framework of the gauge gravitation theory, it may be also applied to the affine-metric gravitation theory and

⁵The connection

$$\tilde{K}^{ab}{}_\alpha = A^{abc}_\mu (\partial_\alpha h_c^\mu + K^\mu_{\nu\lambda} h_c^\nu)\tag{5.471}$$

is not the connection

$$K^k{}_{m\lambda} = h_\mu^k (\partial_\lambda h_m^\mu + K^\mu_{\nu\lambda} h_m^\nu) = K^{ab}{}_\alpha (\eta_{am} \delta_b^k - \eta_{bm} \delta_a^k)$$

written with respect to the reference frame $h^a = h_a^\alpha dx^\alpha$, but there is the relation

$$\tilde{K}^{ab}{}_\alpha = \frac{1}{2} (K^{ab}{}_\alpha - K^{ba}{}_\alpha).\tag{5.472}$$

If K is a Lorentz connection A_h , then the connection \tilde{K} given by (5.471) is consistent with K itself.

the conventional Einstein's gravitation theory. We are not concerned here with the general problem of equivalence of metric, affine and affine-metric theories of gravity [Ferraris and Kijowski (1982)]. At the same time, when K is the *Levi-Civita connection* of h , the Lagrangian density of fermion fields which uses the covariant derivative (5.470) becomes that in the Einstein's gravitation theory. It follows that the configuration space of metric (or tetrad) gravitational fields and general linear connections may play the role of the universal configuration space of realistic gravitational models. In particular, one then can think of the generalized Komar superpotential as being the universal superpotential of energy-momentum of gravity [Giachetta and Sardanashevily (1995)].

We follow [Giachetta and Sardanashevily (1997)] in the geometrical approach to field theory when classical fields are described by global sections of a fibre bundle $Y \rightarrow X$ over a smooth world space-time manifold X . Their dynamics is phrased in terms of jet spaces [Sardanashevily (1993); Saunders (1989)]. Recall that a k th-order differential operator on sections of a fibre bundle $Y \rightarrow X$ is defined to be a bundle morphism of the jet bundle $J^k(X, Y) \rightarrow X$ to a vector bundle over X .

In particular, given bundle coordinates (x^μ, y^i) of a fibre bundle $Y \rightarrow X$, the 1-jet space $J^1(X, Y)$ of Y has the adapted coordinates (x^μ, y^i, y_μ^i) , where $y_\mu^i(j_x^1 s) = \partial_\mu s^i(x)$.

There is the 1-1 correspondence between the connections on the fibre bundle $Y \rightarrow X$ and the global sections $\Gamma = dx^\alpha \otimes (\partial_\alpha + \Gamma_\alpha^i \partial_i)$ of the affine jet bundle $J^1(X, Y) \rightarrow Y$. Every connection Γ on $Y \rightarrow X$ induces the first-order differential operator on Y ,

$$D_\Gamma : J^1(X, Y) \longrightarrow T^*X \otimes VY, \quad D_\Gamma = (y_\alpha^i - \Gamma_\alpha^i) dx^\alpha \otimes \partial_i,$$

which is called the *covariant differential* relative to the connection Γ .

Recall that in the *first-order Lagrangian formalism*, the 1-jet space $J^1(X, Y)$ of Y plays the role of the finite-dimensional configuration space of fields represented by sections s of a bundle $Y \rightarrow X$. A first-order Lagrangian density $L : J^1(X, Y) \rightarrow \wedge^n T^*X$ is defined to be a *horizontal density* $L = \mathcal{L}(x^\mu, y^i, y_\mu^i) \omega$ on the jet bundle $J^1(X, Y) \rightarrow X$, where $\omega = dx^1 \wedge \dots \wedge dx^n$, ($n = \dim X$). Since the jet bundle $J^1(X, Y) \rightarrow Y$ is affine, every polynomial Lagrangian density of field theory factors through $L : J^1(X, Y) \xrightarrow{D} T^*X \otimes VY \rightarrow \wedge^n T^*X$, where D is the covariant differential on Y , and VY is the vertical tangent bundle of Y .

Let us consider the gauge theory of gravity and fermion fields. By X

is further meant an oriented 4D world manifold which satisfies the well-known topological conditions in order that gravitational fields and spinor structure can exist on X . To summarize these conditions, we assume that X is not compact and that the tangent bundle of X is trivial [Giachetta and Sardanashevily (1997)].

Let LX be the principal bundle of oriented linear frames in tangent spaces to X . In gravitation theory, its structure group $GL^+(4, \mathbb{R})$ is reduced to the connected Lorentz group $L = SO(1, 3)$. It means that there exists a reduced subbundle $L^h X$ of LX whose structure group is L . In accordance with the well-known Theorem, there is the 1–1 correspondence between the reduced L subbundles $L^h X$ of LX and the global sections h of the quotient bundle

$$\Sigma = LX/L \rightarrow X. \quad (5.473)$$

These sections h describe gravitational fields on X , for the bundle (5.473) is the 2–folder covering of the bundle of pseudo–Riemannian metrics on X .

Given a section h of Σ , let Ψ^h be an atlas of LX such that the corresponding local sections z_ξ^h of LX take their values into $L^h X$. With respect to Ψ^h and a holonomic atlas $\Psi^T = \{\psi_\xi^T\}$ of LX , a gravitational field h can be represented by a family of GL_4 –valued tetrad functions

$$h_\xi = \psi_\xi^T \circ z_\xi^h, \quad dx^\alpha = h_a^\alpha(x) h^a. \quad (5.474)$$

By the Lorentz connections A_h associated with a gravitational field h are meant the principal connections on the reduced subbundle $L^h X$ of LX . They give rise to principal connections on LX and to spinor connections on the L_s –lift P_h of $L^h X$.

Given a Minkowski space M , let $Cl_{1,3}$ be the *complex Clifford algebra*⁶ generated by elements of M . A spinor space V is defined to be a minimal

⁶Recall that *Clifford algebras* are a type of associative algebra, named after English geometer W. Clifford. They can be thought of as one of the possible generalizations of the complex numbers and quaternions. The theory of Clifford algebras is intimately connected with the theory of quadratic forms and orthogonal transformations. The most important Clifford algebras are those over \mathbb{R} and \mathbb{C} equipped with *nondegenerate quadratic forms*. Recall that every nondegenerate quadratic form on a finite–dimensional real vector space is equivalent to the standard diagonal form $Q(x) = x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_{p+q}^2$, where $n = p + q$ is the dimension of the vector space. The pair of integers (p, q) is called the *signature* of the quadratic form. Similarly, one can define Clifford algebras on complex vector spaces. Every nondegenerate quadratic form on a complex vector space is equivalent to the standard diagonal form $Q(z) = z_1^2 + z_2^2 + \cdots + z_n^2$, so there is essentially only one Clifford algebra in each dimension. One can show that the complex Clifford algebra may be obtained as the complexification of the real one.

left ideal of $Cl_{1,3}$ on which this algebra acts on the left. We have the representation $\gamma : M \otimes V \rightarrow V$ of elements of the Minkowski space $M \subset Cl_{1,3}$ by Dirac's matrices γ on V .

Let us consider a bundle of complex Clifford algebras $Cl_{1,3}$ over X whose structure group is the Clifford group of invertible elements of $Cl_{1,3}$. Its subbundles are both a spinor bundle $S_M \rightarrow X$ and the bundle $Y_M \rightarrow X$ of Minkowski spaces of generating elements of $Cl_{1,3}$. To describe Dirac fermion fields on a world manifold X , one must require Y_M to be isomorphic to the cotangent bundle T^*X of X . It takes place if there exists a reduced L subbundle $L^h X$ such that

$$Y_M = (L^h X \times M)/L.$$

Then, the spinor bundle

$$S_M = S_h = (P_h \times V)/L_s \quad (5.475)$$

is associated with the L_s -lift P_h of $L^h X$. In this case, there exists the representation

$$\gamma_h : T^*X \otimes S_h = (P_h \times (M \otimes V))/L_s \rightarrow (P_h \times \gamma(M \otimes V))/L_s = S_h \quad (5.476)$$

of cotangent vectors to a world manifold X by Dirac's γ -matrices on elements of the spinor bundle S_h . As a shorthand, one can write

$$\hat{d}x^\alpha = \gamma_h(dx^\alpha) = h_a^\alpha(x)\gamma^a.$$

Given the representation (5.476), we shall say that sections of the spinor bundle S_h describe Dirac fermion fields in the presence of the gravitational field h . Let a principal connection on S_h be given by

$$A_h = dx^\alpha \otimes (\partial_\alpha + \frac{1}{2}A^{ab}{}_\alpha I_{ab}{}^A{}_B \psi^B \partial_A).$$

Given the corresponding covariant differential D and the representation γ_h (5.476), one can construct the Dirac operator on the spinor bundle S_h , as

$$\begin{aligned} \mathcal{D}_h &= \gamma_h \circ D : J^1 S_h \rightarrow T^*X \otimes V S_h \rightarrow V S_h, \\ \dot{y}^A \circ \mathcal{D}_h &= h_a^\alpha \gamma^a A_B (y_\alpha^B - \frac{1}{2}A^{ab}{}_\alpha I_{ab}{}^A{}_B y^B). \end{aligned} \quad (5.477)$$

Different gravitational fields h and h' define nonequivalent representations γ_h and $\gamma_{h'}$. It follows that a Dirac fermion field must be regarded only in a pair with a certain gravitational field. There is the 1-1 correspondence between these pairs and sections of the composite spinor bundle (5.469).

Recall that we have a *composite bundle*

$$Y \rightarrow \Sigma \rightarrow X \quad (5.478)$$

of a bundle $Y \rightarrow X$ denoted by Y_Σ and a bundle $\Sigma \rightarrow X$. It is coordinated by $(x^\alpha, \sigma^m, y^i)$ where (x^μ, σ^m) are coordinates of Σ and y^i are the fibre coordinates of Y_Σ . We further assume that Σ has a global section.

The application of composite bundles to field theory is founded on the following [Sardanashvily (1992)]. Given a global section h of Σ , the restriction Y_h of Y_Σ to $h(X)$ is a subbundle of $Y \rightarrow X$. There is the 1–1 correspondence between the global sections s_h of Y_h and the global sections of the composite bundle (5.478) which cover h . Therefore, one can think of sections s_h of Y_h as describing fermion fields in the presence of a background parameter field h , whereas sections of the composite bundle Y describe all the pairs (s_h, h) . The configuration space of these pairs is the 1-jet space $J^1(X, Y)$ of the composite bundle Y .

Every connection

$$A_\Sigma = dx^\alpha \otimes (\partial_\alpha + \tilde{A}_\alpha^i \partial_i) + d\sigma^m \otimes (\partial_m + A_m^i \partial_i)$$

on the bundle Y_Σ induces the *horizontal splitting*

$$VY = VY_\Sigma \oplus (Y \times V\Sigma), \quad \text{locally given by} \\ \dot{y}^i \partial_i + \dot{\sigma}^m \partial_m = (\dot{y}^i - A_m^i \dot{\sigma}^m) \partial_i + \dot{\sigma}^m (\partial_m + A_m^i \partial_i).$$

Using this splitting, one can construct the first-order differential operator (5.41) on the composite bundle Y , namely

$$\tilde{D} : J^1(X, Y) \rightarrow T^*X \otimes VY_\Sigma, \quad \tilde{D} = dx^\alpha \otimes (y_\alpha^i - \tilde{A}_\alpha^i - A_m^i \sigma_\alpha^m) \partial_i. \quad (5.479)$$

This operator possess the following property. Given a global section h of Σ , let Γ be a connection on Σ whose integral section is h , that is, $\Gamma \circ h = j^1 h$. Note that the differential (5.479) restricted to $J^1(X, Y)_h \subset J^1(X, Y)$ becomes the familiar covariant differential relative to the connection on Y_h ,

$$A_h = dx^\alpha \otimes [\partial_\alpha + (A_m^i \partial_\alpha h^m + \tilde{A}_\alpha^i) \partial_i].$$

Thus, it is \tilde{D} that we may use in order to construct a Lagrangian density

$$L : J^1(X, Y) \xrightarrow{\tilde{D}} T^*X \otimes VY_\Sigma \rightarrow \wedge^n T^*X$$

for sections of the composite bundle Y .

In particular, in gravitation theory, we have the composite bundle $LX \rightarrow \Sigma \rightarrow X$, where Σ is the quotient bundle (5.473) and $LX_\Sigma =$

$LX \rightarrow \Sigma$ is the L -principal bundle. Let P_Σ be the L_s -principal lift of LX_Σ such that $P_\Sigma/L_s = \Sigma$ and $LX_\Sigma = r(P_\Sigma)$. In particular, there is the imbedding of the L_s -lift P_h of L^hX onto the restriction of P_Σ to $h(X)$ [Giachetta and Sardanashvily (1997)].

Let us consider the composite spinor bundle (5.469) where $S_\Sigma = (P_\Sigma \times V)/L_s$ is associated with the L_s -principal bundle P_Σ . Note that, given a global section h of Σ , the restriction S_Σ to $h(X)$ is the spinor bundle S_h (5.475) whose sections describe Dirac fermion fields in the presence of the gravitational field h .

Let us give the principal bundle LX with a holonomic atlas $\{\psi_\xi^T, U_\xi\}$ and the principal bundles P_Σ and LX_Σ with associated atlases $\{z_\epsilon^s, U_\epsilon\}$ and $\{z_\epsilon = r \circ z_\epsilon^s\}$. With respect to these atlases, the composite spinor bundle is equipped with the bundle coordinates $(x^\alpha, \sigma_a^\mu, \psi^A)$ where (x^α, σ_a^μ) are coordinates of the bundle Σ such that σ_a^μ are the matrix components of the group element $(\psi_\xi^T \circ z_\epsilon)(\sigma)$, $\sigma \in U_\epsilon$, $\pi_{\Sigma X}(\sigma) \in U_\xi$. Given a section h of Σ , we have $(\sigma_a^\alpha \circ h)(x) = h_a^\alpha(x)$, where $h_a^\alpha(x)$ are the tetrad functions (5.474).

Let us consider the bundle of Minkowski spaces $(LX \times M)/L \rightarrow \Sigma$ associated with the L -principal bundle LX_Σ . Since LX_Σ is trivial, it is isomorphic to the pull-back $\Sigma \times T^*X$ which we denote by the same symbol T^*X . Then, one can define the bundle morphism γ_Σ over Σ , given by

$$\gamma_\Sigma : T^*X \otimes S_\Sigma \rightarrow S_\Sigma, \quad \hat{d}x^\alpha = \gamma_\Sigma(dx^\alpha) = \sigma_a^\alpha \gamma^a. \quad (5.480)$$

When restricted to $h(X) \subset \Sigma$, the map (5.480) becomes the morphism γ_h (5.476). We use this morphism in order to construct the total Dirac operator on the composite spinor bundle S (5.469).

Let

$$\tilde{A} = dx^\alpha \otimes (\partial_\alpha + \tilde{A}_\alpha^B \partial_B) + d\sigma_a^\mu \otimes (\partial_\mu^a + A^{Ba}_\mu \partial_B)$$

be a principal connection on the bundle S_Σ and \tilde{D} the corresponding differential (5.479). We have the first-order differential operator on S , given by

$$\begin{aligned} \mathcal{D} &= \gamma_\Sigma \circ \tilde{D} : J^1S \rightarrow T^*X \otimes VS_\Sigma \rightarrow VS_\Sigma, \\ \dot{\psi}^A \circ \mathcal{D} &= \sigma_a^\alpha \gamma^{aA}_B (\psi_\alpha^B - \tilde{A}_\alpha^B - A^{Ba}_\mu \sigma_a^\mu). \end{aligned}$$

One can think of it as being the total Dirac operator since, for every section h , the restriction of \mathcal{D} to $J^1S_h \subset J^1S$ becomes the Dirac operator \mathcal{D}_h

(5.477) relative to the connection on the bundle S_h , given by

$$A_h = dx^\alpha \otimes [\partial_\alpha + (\tilde{A}_\alpha^B + A^{Ba}_\mu \partial_\alpha h^\mu_a) \partial_B].$$

In order to construct the differential \tilde{D} (5.479) on $J^1(X, S)$ in explicit form, let us consider the principal connection on the bundle LX_Σ which is given by the local connection form

$$\tilde{A} = (\tilde{A}^{ab}_\mu dx^\mu + A^{abc}_\mu d\sigma^\mu_c) \otimes I_{ab}, \quad (5.481)$$

$$\begin{aligned} \tilde{A}^{ab}_\mu &= \frac{1}{2} K^\nu_{\lambda\mu} \sigma^\alpha_c (\eta^{ca} \sigma^\mu_\nu - \eta^{cb} \sigma^\mu_\nu), \\ A^{abc}_\mu &= \frac{1}{2} (\eta^{ca} \sigma^\mu_\mu - \eta^{cb} \sigma^\mu_\mu), \end{aligned} \quad (5.482)$$

where K is a general linear connection on TX and (5.482) corresponds to the canonical left-invariant connection on the bundle $GL^+(4, \mathbb{R}) \rightarrow GL^+(4, \mathbb{R})/L$.

Therefore, the differential \tilde{D} relative to the connection (5.481) reads

$$\tilde{D} = dx^\alpha \otimes [\partial_\alpha - \frac{1}{2} A^{abc}_\mu (\sigma^\mu_{c\lambda} + K^\mu_{\nu\lambda} \sigma^\nu_c) I_{ab}^A \psi^B \partial_A]. \quad (5.483)$$

Given a section h , the connection \tilde{A} (5.481) is reduced to the Lorentz connection \tilde{K} (5.471) on $L^h X$, and the differential (5.483) leads to the covariant derivatives of fermion fields (5.470). We will use the differential (5.483) in order to construct a Lagrangian density of Dirac fermion fields. Their Lagrangian density is defined on the configuration space $J^1(X, S \oplus S^+)$ coordinated by $(x^\mu, \sigma^\mu_a, \psi^A, \psi^+_A, \sigma^\mu_{a\lambda}, \psi^A_\alpha, \psi^+_{A\lambda})$. It reads

$$\begin{aligned} L_\psi &= \left\{ \frac{i}{2} [\psi^+_A (\gamma^0 \gamma^\alpha)^A_B (\psi^B_\alpha - \frac{1}{2} A^{abc}_\mu (\sigma^\mu_{c\lambda} + K^\mu_{\nu\lambda} \sigma^\nu_c) I_{ab}^B \psi^C) \right. \\ &\quad - (\psi^+_{A\lambda} - \frac{1}{2} A^{abc}_\mu (\sigma^\mu_{c\lambda} + K^\mu_{\nu\lambda} \sigma^\nu_c) \psi^+_C I_{ab}^C \psi^A) (\gamma^0 \gamma^\alpha)^A_B \psi^B] \\ &\quad \left. - m \psi^+_A (\gamma^0)^A_B \psi^B \right\} \sigma^{-1} \omega, \end{aligned} \quad (5.484)$$

where $\gamma^\mu = \sigma^\mu_a \gamma^a$, and $\sigma = \det(\sigma^\mu_a)$, while $\psi^+_A (\gamma^0)^A_B \psi^B$ is the Lorentz-invariant fibre metric in the bundle $S \oplus S^*$ [Crawford (1991)].

One can show that

$$\frac{\partial \mathcal{L}_\psi}{\partial K^\mu_{\nu\lambda}} + \frac{\partial \mathcal{L}_\psi}{\partial K^\mu_{\lambda\nu}} = 0.$$

Hence, the Lagrangian density (5.484) depends on the torsion of the general linear connection K only. In particular, it follows that, if K is the

Levi-Civita connection of a gravitational field $h(x)$, after the substitution $\sigma_c^\nu = h_c^\nu(x)$, the Lagrangian density (5.484) becomes the familiar Lagrangian density of fermion fields in the Einstein's gravitation theory.

5.12.3 *Hawking–Penrose Quantum Gravity and Black Holes*

In their search for quantum gravity, S. Hawking and R. Penrose use the *straightforward* application of quantum theory to general relativity [Hawking and Israel (1979); Penrose (1989); Hawking and Penrose (1996)], rather than following the more fashioned string theory approach (described below).

According to Hawking, “Einstein’s general relativity is a beautiful theory that agrees with every observation that has been made so far. It might require modifications on the Planck scale, and it might be only a *low energy approximation* to some more fundamental theory, like e.g., superstring theory, but it will not affect many of the predictions that can be get from gravity...” [Hawking and Israel (1979)].

Space–Time Manifold, Gravity, Black Holes and Big Bang

The crucial technique for investigating *Hawking–Penrose singularities and black holes*, has been the study of the global causal structure of space–time [Hawking and Israel (1979)]. Define $I^+(p)$ to be the set of all points of the space–time manifold M that can be reached from the point p by future directed time like curves. One can think of $I^+(p)$ as the set of all events that can be influenced by what happens at p . One now considers the boundary $\dot{I}^+(S)$ of the future of a set S . It is easy to see that this boundary cannot be *time-like*. For in that case, a point q just outside the boundary would be to the future of a point p just inside. Nor can the boundary of the future be *space-like*, except at the set S itself. For in that case every past directed curve from a point q , just to the future of the boundary, would cross the boundary and leave the future of S . That would be a contradiction with the fact that q is in the future of S . Therefore, the boundary of the future is null apart from at S itself.

To show that each generator of the boundary of the future has a past end point on the set, one has to impose some global condition on the causal structure. The strongest and physically most important condition is that of

global space-time hyperbolicity.⁷ The significance of global hyperbolicity for singularity theorems stems from the following [Hawking and Israel (1979); Hawking and Penrose (1996)]. Let U be globally hyperbolic and let p and q be points of U that can be joined by a time like or null curve. Then there is a time-like or null-geodesic between p and q which maximizes the length of time like or null curves from p to q . The method of proof is to show the space of all time like or null curves from p to q is compact in a certain topology. One then shows that the length of the curve is an upper semi-continuous function on this space. It must therefore attain its maximum and the curve of maximum length will be a geodesic because otherwise a small variation will give a longer curve.

One can now consider the second variation of the length of a geodesic γ . One can show that γ can be varied to a longer curve if there is an infinitesimally neighboring geodesic from p which intersects γ again at a point r between p and q . The point r is said to be conjugate to p . One can illustrate this by considering two points p and q on the surface of the Earth. Without loss of generality one can take p to be at the north pole. Because the Earth has a positive definite metric rather than a Lorentzian one, there is a geodesic of minimal length, rather than a geodesic of maximum length. This minimal geodesic will be a line of longitude running from the north pole to the point q . But there will be another geodesic from p to q which runs down the back from the north pole to the south pole and then up to q . This geodesic contains a point conjugate to p at the south pole where all the geodesics from p intersect. Both geodesics from p to q are stationary points of the length under a small variation. But now in a positive definite metric the second variation of a geodesic containing a conjugate point can give a shorter curve from p to q . Thus, on the Earth, the geodesic that goes down to the south pole and then comes up is not the shortest curve from p to q .

The reason one gets conjugate points in space-time is that gravity is an attractive force. It therefore curves space-time in such a way that neighboring geodesics are bent towards each other rather than away. One can

⁷Recall that an open set U is said to be globally hyperbolic if:

- (1) For every pair of points p and q in U the intersection of the future of p and the past of q has compact closure. In other words, it is a bounded diamond shaped region.
- (2) Strong causality holds on U . That is, there are no closed or almost closed time-like curves contained in U .

see this from the *Newman–Penrose equation*

$$\frac{d\rho}{dv} = \rho^2 + \sigma^{ij}\sigma_{ij} + \frac{1}{n}R_{\alpha\beta}l^\alpha l^\beta, \quad (\alpha, \beta = 0, 1, 2, 3)$$

where $n = 2$ for null geodesics and $n = 3$ for time-like geodesics. Here v is an affine parameter along a congruence of geodesics, with tangent vector l^α which are hypersurface orthogonal. The quantity ρ is the average rate of convergence of the geodesics, while σ measures the shear. The term $R_{\alpha\beta}l^\alpha l^\beta$ gives the direct gravitational effect of the matter on the convergence of the geodesics. By the *Einstein equation* (3.1), it will be non-negative for any null vector l^α if the matter obeys the so-called *weak energy condition*, which says that the energy density T_{00} is non-negative in any frame, i.e.,

$$T_{\alpha\beta}v^\alpha v^\beta \geq 0, \quad (5.485)$$

for any time-like vector v^α , is obeyed by the classical SEM-tensor of any reasonable matter [Hawking and Israel (1979); Hawking and Penrose (1996)].

Suppose the weak energy condition holds, and that the null geodesics from a point p begin to converge again and that ρ has the positive value ρ_0 . Then the Newman–Penrose equation would imply that the convergence ρ would become infinite at a point q within an affine parameter distance $\frac{1}{\rho_0}$ if the null geodesic can be extended that far. If $\rho = \rho_0$ at $v = v_0$ then $\rho \geq \frac{1}{\rho^{-1} + v_0 - v}$. Thus there is a conjugate point before $v = v_0 + \rho^{-1}$.

Infinitesimally neighboring null geodesics from p will intersect at q . This means the point q will be conjugate to p along the null geodesic γ joining them. For points on γ beyond the conjugate point q there will be a variation of γ that gives a time like curve from p . Thus γ cannot lie in the boundary of the future of p beyond the conjugate point q . So γ will have a future end point as a generator of the boundary of the future of p .

The situation with time-like geodesics is similar, except that the *strong energy condition* [Hawking and Israel (1979); Hawking and Penrose (1996)],

$$T_{\alpha\beta}v^\alpha v^\beta \geq \frac{1}{2}v^\alpha v_\alpha T, \quad (5.486)$$

that is required to make $R_{\alpha\beta}l^\alpha l^\beta$ non-negative for every time like vector l^α , is rather stronger than the weak energy condition (5.485). However, it is still physically reasonable, at least in an averaged sense, in classical theory. If the strong energy condition holds, and the time like geodesics from p begin converging again, then there will be a point q conjugate to p .

Finally there is the *generic energy condition*, which says:

- (1) The strong energy condition holds.
- (2) Every time-like or null geodesic has a point where $l_{[a}R_{b]cd[e}l_{f]}l^cl^d \neq 0$.

One normally thinks of a space-time singularity as a region in which the curvature becomes unboundedly large. However, the trouble with this definition is that one could simply leave out the singular points and say that the remaining manifold was the whole of space-time. It is therefore better to define space-time as the maximal manifold on which the metric is suitably smooth. One can then recognize the occurrence of singularities by the existence of incomplete geodesics that cannot be extended to infinite values of the affine parameter.

Hawking–Penrose Singularity Theorems

Hawking–Penrose Singularity is defined as follows [Hawking and Israel (1979); Penrose (1989); Hawking and Penrose (1996)]:

A space-time manifold is singular if it is time-like or null geodesically incomplete but cannot be embedded in a larger space-time manifold.

This definition reflects the most objectionable feature of singularities, that there can be particles whose history has a beginning or end at a finite time. There are examples in which geodesic incompleteness can occur with the curvature remaining bounded, but it is thought that generically the curvature will diverge along incomplete geodesics. This is important if one is to appeal to quantum effects to solve the problems raised by singularities in classical general relativity.

Singularity Theorems include:

- (1) Energy condition (i.e., weak (5.485), strong (5.486), or generic (5.12.3)).
- (2) Condition on global structure (e.g., there should not be any closed time-like curves).
- (3) Gravity strong enough to trap a region (so that nothing could escape).

The various singularity theorems show that space-time must be time like or null geodesically incomplete if different combinations of the three kinds of conditions hold. One can weaken one condition if one assumes stronger versions of the other two. The Hawking–Penrose Singularity theorems have the generic energy condition, the strongest of the three energy conditions. The global condition is fairly weak, that there should be no closed time like

curves. And the no escape condition is the most general, that there should be either a trapped surface or a closed space like three surface.

The theorems predict singularities in two situations. One is in the future in the gravitational collapse of stars and other massive bodies. Such singularities would be an end of time, at least for particles moving on the incomplete geodesics. The other situation in which singularities are predicted is in the past at the beginning of the present expansion of the universe.

The prediction of singularities means that classical general relativity is not a complete theory. Because the singular points have to be cut out of the space-time manifold one cannot define the field equations there and cannot predict what will come out of a singularity. With the singularity in the past the only way to deal with this problem seems to be to appeal to quantum gravity. But the singularities that are predicted in the future seem to have a property that Penrose has called, Cosmic Censorship. That is they conveniently occur in places like black holes that are hidden from external observers. So any break down of predictability that may occur at these singularities will not affect what happens in the outside world, at least not according to classical theory.

Hawking Cosmic Censorship Hypothesis says: "Nature abhors a naked singularity" [Hawking and Israel (1979); Hawking and Penrose (1996)]. However, there is unpredictability in the quantum theory. This is related to the fact that gravitational fields can have intrinsic entropy which is not just the result of coarse graining. *Gravitational entropy*, and the fact that time has a beginning and may have an end, are the two main themes of Hawking's research, because they are the ways in which gravity is distinctly different from other physical fields.

The fact that gravity has a quantity that behaves like entropy was first noticed in the purely classical theory. It depends on Penrose's Cosmic Censorship Conjecture. This is unproved but is believed to be true for suitably general initial data and state equations.

One makes the approximation of treating the region around a collapsing star as asymptotically flat. Then, as Penrose showed, one can conformally embed the space-time manifold M in a manifold with boundary \bar{M} . The boundary ∂M will be a null surface and will consist of two components, future and past null infinity, called \mathcal{I}^+ and \mathcal{I}^- . One says that *weak* Cosmic Censorship holds if two conditions are satisfied. First, it is assumed that the null geodesic generators of \mathcal{I}^+ are complete in a certain conformal metric. This implies that observers far from the collapse live to an old age and are not wiped out by a thunderbolt singularity sent out from the collapsing

star. Second, it is assumed that the past of \mathcal{I}^+ is globally hyperbolic. This means there are no naked singularities that can be seen from large distances. Penrose has also a stronger form of Cosmic Censorship which assumes that the whole space-time is globally hyperbolic.

Weak Cosmic Censorship Hypothesis reads:

- (1) \mathcal{I}^+ and \mathcal{I}^- are complete.
- (2) $I^-(\mathcal{I}^+)$ is globally hyperbolic.

If weak Cosmic Censorship holds, the singularities that are predicted to occur in gravitational collapse cannot be visible from \mathcal{I}^+ . This means that there must be a region of space-time that is not in the past of \mathcal{I}^+ . This region is said to be a **black hole** because no light or anything else can escape from it to infinity. The boundary of the black hole region is called the *event horizon*. Because it is also the boundary of the past of \mathcal{I}^+ the event horizon will be generated by null-geodesic segments that may have past end points but don't have any future end points. It then follows that if the weak energy condition holds the generators of the horizon cannot be converging. For if they were they would intersect each other within a finite distance [Hawking and Israel (1979); Penrose (1989); Hawking and Penrose (1996)].

This implies that the area of a cross section of the event horizon can never decrease with time and in general will increase. Moreover if two black holes collide and merge together the area of the final black hole will be greater than the sum of the areas of the original black holes. This is very similar to the behavior of entropy according to the *Second Law of Thermodynamics*:⁸.

Second Law of Black Hole Mechanics: $\delta\mathcal{A} \geq 0$.

Second Law of Thermodynamics: $\delta S \geq 0$.

The similarity with thermodynamics is increased by what is called the First Law of Black Hole Mechanics, which relates the change in mass of a black hole to the change in the area of the event horizon and the change in its angular momentum and electric charge. One can compare this to the First Law of Thermodynamics which gives the change in internal energy in terms of the change in entropy and the external work done on the system [Hawking and Israel (1979); Hawking and Penrose (1996)]:

First Law of Black Hole Mechanics: $\delta E = \frac{\kappa}{8\pi} \delta\mathcal{A} + \Omega\delta J + \Phi\delta Q$.

First Law of Thermodynamics: $\delta E = T\delta S + P\delta V$.

⁸Recall that Second Law of Thermodynamics states: Entropy can never decrease and the entropy of a total system is greater than the sum of its constituent parts

One sees that if the area \mathcal{A} of the event horizon is analogous to entropy S then the quantity analogous to temperature is what is called the surface gravity of the black hole κ . This is a measure of the strength of the gravitational field on the event horizon. The similarity with thermodynamics is further increased by the so-called Zeroth Law of Black Hole Mechanics: the surface gravity is the same everywhere on the event horizon of a time independent black hole [Hawking and Israel (1979)].

Zeroth Law of Black Hole Mechanics:

κ is the same everywhere on the horizon of a time independent black hole.

Zeroth Law of Thermodynamics:

T is the same everywhere for a system in thermal equilibrium.

Encouraged by these similarities Bekenstein proposed that some multiple of the area of the event horizon actually was the entropy of a black hole. He suggested a generalized Second Law: the sum of this black hole entropy and the entropy of matter outside black holes would never decrease (see [Strominger and Vafa (1996)]).

Generalized Second Law: $\delta(S + c\mathcal{A}) \geq 0$.

However, this proposal was not consistent. If black holes have an entropy proportional to horizon area \mathcal{A} they should also have a non zero temperature proportional to surface gravity.

Path-Integral Model for Black Holes

Recall that the fact that gravity is attractive means that it will tend to draw the matter in the universe together to form objects like stars and galaxies. These can support themselves for a time against further contraction by thermal pressure, in the case of stars, or by rotation and internal motions, in the case of galaxies. However, eventually the heat or the angular momentum will be carried away and the object will begin to shrink. If the mass is less than about one and a half times that of the Sun the contraction can be stopped by the degeneracy pressure of electrons or neutrons. The object will settle down to be a white dwarf or a neutron star respectively. However, if the mass is greater than this limit there is nothing that can hold it up and stop it continuing to contract. Once it has shrunk to a certain critical size the gravitational field at its surface will be so strong that the light cones will be bent inward [Hawking and Israel (1979); Hawking and Penrose (1996)].

If the Cosmic Censorship Conjecture is correct the trapped surface and the singularity it predicts cannot be visible from far away. Thus there must

be a region of space-time from which it is not possible to escape to infinity. This region is said to be a black hole. Its boundary is called the event horizon and it is a null surface formed by the light rays that just fail to get away to infinity. As we saw in the last subsection, the area \mathcal{A} of a cross section of the event horizon can never decrease, at least in the classical theory. This, and perturbation calculations of spherical collapse, suggest that black holes will settle down to a stationary state.

Recall that the *Schwarzschild metric form*, given by

$$ds^2 = -\left(1 - \frac{2M}{r}\right)dt^2 + \left(1 - \frac{2M}{r}\right)^{-1}dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2),$$

represents the gravitational field that a black hole would settle down to if it were non rotating. In the usual r and t coordinates there is an apparent singularity at the Schwarzschild radius $r = 2M$. However, this is just caused by a bad choice of coordinates. One can choose other coordinates in which the metric is regular there.

Now, if one performs the *Wick rotation*, $t = i\tau$, one gets a positive definite metric, usually called Euclidean even though they may be curved. In the *Euclidean-Schwarzschild metric*

$$ds^2 = x^2 \left(\frac{d\tau}{4M}\right)^2 + \left(\frac{r^2}{4M^2}\right)^2 dx^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)$$

there is again an apparent singularity at $r = 2M$. However, one can define a new radial coordinate x to be $4M(1 - 2Mr^{-1})^{\frac{1}{2}}$.

The metric in the $x - \tau$ plane then becomes like the origin of polar coordinates if one identifies the coordinate τ with period $8\pi M$. Similarly, other Euclidean black hole metrics will have apparent singularities on their horizons which can be removed by identifying the imaginary time coordinate with period $\frac{2\pi}{\kappa}$.

To see the significance of having imaginary time identified with some period β , let us consider the *amplitude* to go from some field configuration ϕ_1 on the surface t_1 to a configuration ϕ_2 on the surface t_2 . This will be given by the matrix element of $e^{iH(t_2-t_1)}$. However, one can also represent this amplitude as a *path integral* over all fields ϕ between t_1 and t_2 which agree with the given fields ϕ_1 and ϕ_2 on the two surfaces,

$$\langle \phi_2, t_2 | \phi_1, t_1 \rangle = \langle \phi_2 | \exp(-iH(t_2 - t_1)) | \phi_1 \rangle = \int D[\phi] \exp(iA[\phi]).$$

One now chooses the time separation $(t_2 - t_1)$ to be pure imaginary and equal to β . One also puts the initial field ϕ_1 equal to the final field ϕ_2 and

sums over a complete basis of states ϕ_n . On the left one has the expectation value of $e^{-\beta H}$ summed over all states. This is just the *thermodynamic partition function* Z at the temperature $T = \beta^{-1}$,

$$Z = \sum \langle \phi_n | \exp(-\beta H) | \phi_n \rangle = \int D[\phi] \exp(-A[\phi]). \quad (5.487)$$

On the r.h.s. of this equation one has a *path integral* (see chapter 6). One puts $\phi_1 = \phi_2$ and sums over all field configurations ϕ_n . This means that effectively one is doing the path integral over all fields ϕ on a space-time that is identified periodically in the imaginary time direction with period β . Thus the partition function for the field ϕ at temperature T is given by a path integral over all fields on a Euclidean space-time. This space-time is periodic in the imaginary time direction with period $\beta = T^{-1}$ [Hawking and Israel (1979); Hawking and Penrose (1996)].

If one calculates the path integral in flat space-time identified with period β in the imaginary time direction one gets the usual result for the partition function of black body radiation. However, as we have just seen, the Euclidean-Schwarzschild solution is also periodic in imaginary time with period $\frac{2\pi}{\kappa}$. This means that fields on the Schwarzschild background will behave as if they were in a thermal state with temperature $\frac{\kappa}{2\pi}$.

The periodicity in imaginary time explained why the messy calculation of frequency mixing led to radiation that was exactly thermal. However, this derivation avoided the problem of the very high frequencies that take part in the frequency mixing approach. It can also be applied when there are interactions between the quantum fields on the background. The fact that the path integral is on a periodic background implies that all physical quantities like expectation values will be thermal. This would have been very difficult to establish in the frequency mixing approach [Hawking and Israel (1979); Hawking and Penrose (1996)].

One can extend these interactions to include interactions with the gravitational field itself. One starts with a background metric g_0 such as the Euclidean-Schwarzschild metric that is a solution of the classical field equations. One can then expand the action A in a power series in the perturbations δg about g_0 , as

$$A[g] = A[g_0] + A_2(\delta g)^2 + A_3(\delta g)^3 + \dots$$

Here, the linear term vanishes because the background is a solution of the field equations. The quadratic term can be regarded as describing gravitons on the background while the cubic and higher terms describe interactions

between the gravitons. The path integral over the quadratic terms are finite. There are non renormalizable divergences at two loops in pure gravity but these cancel with the fermions in super-gravity theories. It is not known whether super-gravity theories have divergences at three loops or higher because no one has been brave or foolhardy enough to try the calculation. Some recent work indicates that they may be finite to all orders. But even if there are higher loop divergences they will make very little difference except when the background is curved on the scale of the Planck length (10^{-33} cm).

More interesting than the higher order terms is the zeroth order term, the action of the background metric g_0 [Hawking and Israel (1979); Hawking and Penrose (1996)],

$$A = -\frac{1}{16\pi} \int R(-g)^{\frac{1}{2}} d^4x + \frac{1}{8\pi} \int K(\pm h)^{\frac{1}{2}} d^3x.$$

Recall that the usual *Einstein-Hilbert action* for general relativity is the volume integral of the scalar curvature R . This is zero for vacuum solutions so one might think that the action of the Euclidean-Schwarzschild solution was zero. However, there is also a surface term in the action proportional to the integral of K , the trace of the second fundamental form of the boundary surface. When one includes this and subtracts off the surface term for flat space one finds the action of the Euclidean-Schwarzschild metric is $\frac{\beta^2}{16\pi}$ where β is the period in imaginary time at infinity. Thus the dominant contribution to the path integral for the partition function Z given by (5.487), is $e^{\frac{-\beta^2}{16\pi}}$,

$$Z = \sum \exp(-\beta E_n) = \exp\left(-\frac{\beta^2}{16\pi}\right).$$

If one differentiates $\log Z$ with respect to the period β one gets the expectation value of the energy, or in other words, the mass,

$$\langle E \rangle = -\frac{d}{d\beta}(\log Z) = \frac{\beta}{8\pi}.$$

So this gives the mass $M = \frac{\beta}{8\pi}$. This confirms the relation between the mass and the period, or inverse temperature, that we already knew. However, one can go further. By standard thermodynamic arguments, the log of the partition function is equal to minus the free energy F divided by the temperature T , i.e., $\log Z = -\frac{F}{T}$. And the free energy is the mass or energy plus the temperature times the entropy S , i.e., $F = \langle E \rangle + TS$. Putting

all this together one sees that the action of the black hole gives an entropy of $4\pi M^2$,

$$S = \frac{\beta^2}{16\pi} = 4\pi M^2 = \frac{1}{4}\mathcal{A}.$$

This is exactly what is required to make the laws of black holes the same as the laws of thermodynamics [Hawking and Israel (1979); Hawking and Penrose (1996)]. The reason why does one get this intrinsic gravitational entropy which has no parallel in other quantum field theories, is that gravity allows different topologies for the space–time manifold.

In the case we are considering the Euclidean–Schwarzschild solution has a boundary at infinity that has topology $S^2 \times S^1$. The S^2 is a large space like two sphere at infinity and the S^1 corresponds to the imaginary time direction which is identified periodical. One can fill in this boundary with metrics of at least two different topologies. One is the Euclidean–Schwarzschild metric. This has topology $R^2 \times S^2$, that is the Euclidean two plane times a two sphere. The other is $R^3 \times S^1$, the topology of Euclidean flat space periodically identified in the imaginary time direction. These two topologies have different Euler numbers. The *Euler number* of periodically identified flat space is zero, while that of the Euclidean–Schwarzschild solution is two,

$$\text{Total action} = M(\tau_2 - \tau_1).$$

The significance of this is as follows: on the topology of periodically identified flat space one can find a periodic time function τ whose gradient is no where zero and which agrees with the imaginary time coordinate on the boundary at infinity. One can then work out the action of the region between two surfaces τ_1 and τ_2 . There will be two contributions to the action, a volume integral over the matter Lagrangian, plus the Einstein–Hilbert Lagrangian and a surface term. If the solution is time independent the surface term over $\tau = \tau_1$ will cancel with the surface term over $\tau = \tau_2$. Thus the only net contribution to the surface term comes from the boundary at infinity. This gives half the mass times the imaginary time interval $(\tau_2 - \tau_1)$. If the mass is non–zero there must be non–zero matter fields to create the mass. One can show that the volume integral over the matter Lagrangian plus the Einstein–Hilbert Lagrangian also gives $\frac{1}{2}M(\tau_2 - \tau_1)$. Thus the total action is $M(\tau_2 - \tau_1)$. If one puts this contribution to the log of the partition function into the thermodynamic formulae one finds

the expectation value of the energy to be the mass, as one would expect. However, the entropy contributed by the background field will be zero.

However, the situation is different with the Euclidean-Schwarzschild solution, which says:

$$\text{Total action including corner contribution} = M(\tau_2 - \tau_1)$$

$$\text{Total action without corner contribution} = \frac{1}{2}M(\tau_2 - \tau_1)$$

Because the Euler number is two rather than zero one cannot find a time function τ whose gradient is everywhere non-zero. The best one can do is choose the imaginary time coordinate of the Schwarzschild solution. This has a fixed two sphere at the horizon where τ behaves like an angular coordinate. If one now works out the action between two surfaces of constant τ the volume integral vanishes because there are no matter fields and the scalar curvature is zero. The trace K surface term at infinity again gives $\frac{1}{2}M(\tau_2 - \tau_1)$. However there is now another surface term at the horizon where the τ_1 and τ_2 surfaces meet in a corner. One can evaluate this surface term and find that it also is equal to $\frac{1}{2}M(\tau_2 - \tau_1)$. Thus the total action for the region between τ_1 and τ_2 is $M(\tau_2 - \tau_1)$. If one used this action with $\tau_2 - \tau_1 = \beta$ one would find that the entropy was zero. However, when one looks at the action of the Euclidean Schwarzschild solution from a 4-dimensional point of view rather than a $3 + 1$, there is no reason to include a surface term on the horizon because the metric is regular there. Leaving out the surface term on the horizon reduces the action by one quarter the area of the horizon, which is just the intrinsic gravitational entropy of the black hole [Hawking and Israel (1979); Hawking and Penrose (1996)].

Quantum Cosmology

According to Hawking, *cosmology* used to be considered a *pseudo-science* and the preserve of physicists who may have done useful work in their earlier years but who had gone mystic in their dotage. There is a serious objection that cosmology cannot predict anything about the universe unless it makes some assumption about the *initial conditions*. Without such an assumption, all one can say is that things are as they are now because they were as they were at an earlier stage. Yet many people believe that science should be concerned only with the local laws which govern how the

universe evolves in time. They would feel that the boundary conditions for the universe that determine how the universe began were a question for metaphysics or religion rather than science [Hawking and Israel (1979); Hawking and Penrose (1996)].

Hawking–Penrose theorems showed that according to general relativity there should be a singularity in our past. At this singularity the field equations could not be defined. Thus classical general relativity brings about its own downfall: it predicts that it cannot predict the universe. For Hawking this sounds really disturbing: If the laws of physics could break down at the beginning of the universe, why couldn't they break down anywhere. In quantum theory it is a principle that anything can happen if it is not absolutely forbidden. Once one allows that singular histories could take part in the path integral they could occur anywhere and predictability would disappear completely. If the laws of physics break down at singularities, they could break down anywhere.

The only way to have a scientific theory is if the laws of physics hold everywhere including at the beginning of the universe. One can regard this as a triumph for the *Principle of Democracy*: Why should the beginning of the universe be exempt from the laws that apply to other points. If all points are equal one cannot allow some to be more equal than others.

To implement the idea that the laws of physics hold everywhere, one should take the path integral only over non-singular metrics. One knows in the ordinary path integral case that the measure is concentrated on non-differentiable paths. But these are the completion in some suitable topology of the set of smooth paths with well defined action. Similarly, one would expect that the path integral for quantum gravity should be taken over the completion of the space of smooth metrics. What the path integral cannot include is metrics with singularities whose action is not defined.

In the case of black holes we saw that the path integral should be taken over Euclidean, that is, positive definite metrics. This meant that the singularities of black holes, like the Schwarzschild solution, did not appear on the Euclidean metrics which did not go inside the horizon. Instead the horizon was like the origin of polar coordinates. The action of the Euclidean metric was therefore well defined. One could regard this as a quantum version of Cosmic Censorship: the break down of the structure at a singularity should not affect any physical measurement.

It seems, therefore, that the path integral for quantum gravity should be taken over non-singular Euclidean metrics. But what should the boundary conditions be on these metrics. There are two, and only two, nat-

ural choices. The first is metrics that approach the flat Euclidean metric outside a compact set. The second possibility is metrics on manifolds that are compact and without boundary. Therefore, the natural choices for path integral for quantum gravity are [Hawking and Israel (1979); Hawking and Penrose (1996)]: (i) asymptotically Euclidean metrics, and (ii) compact metrics without boundary. The first class of asymptotically Euclidean metrics is appropriate for *scattering* calculations. In these one sends particles in from infinity and observes what comes out again to infinity. All measurements are made at infinity where one has a flat background metric and one can interpret small fluctuations in the fields as particles in the usual way. One doesn't ask what happens in the interaction region in the middle. That is why one does a path integral over all possible histories for the interaction region, that is, over all asymptotically Euclidean metrics. However, in cosmology one is interested in measurements that are made in a finite region rather than at infinity. We are on the inside of the universe not looking in from the outside. To see what difference this makes let us first suppose that the path integral for cosmology is to be taken over all asymptotically Euclidean metrics.

The so-called No Boundary Proposal of Hartle and Hawking reads [Hawking and Israel (1979); Hawking and Penrose (1996)]: The path integral for quantum gravity should be taken over all compact Euclidean metrics. One can paraphrase this as: *the boundary condition of the universe is that it has no boundary*. According to Hawking, this no boundary proposal seems to account for the universe we live in. That is an isotropic and homogeneous expanding universe with small perturbations. We can observe the spectrum and statistics of these perturbations in the fluctuations in the microwave background. The results so far agree with the predictions of the no boundary proposal. It will be a real test of the proposal and the whole Euclidean quantum gravity program when the observations of the microwave background are extended to smaller angular scales.

In order to use the no boundary proposal to make predictions, it is useful to introduce a concept that can describe the state of the universe at one time:

$$\text{Probability of induced metric } h_{ij} \text{ on } \Sigma = \int_{\substack{\text{metrics on } M \text{ that} \\ \text{induce } h_{ij} \text{ on } \Sigma}} d[g] \exp(-A[g]).$$

Consider the probability that the space-time manifold M contains an embedded three dimensional manifold Σ with induced metric h_{ij} . This is given by a path integral over all metrics g_{ab} on M that induce h_{ij}

on Σ . If M is simply-connected, which we will assume, the surface Σ will divide M into two parts M^+ and M^- [Hawking and Israel (1979); Hawking and Penrose (1996)],

Probability of $h_{ij} = \Psi^+(h_{ij}) \times \Psi^-(h_{ij})$, where

$$\Psi^+(h_{ij}) = \int_{\substack{\text{metrics on } M^+ \text{ that} \\ \text{induce } h_{ij} \text{ on } \Sigma}} d[g] \exp(-A[g]).$$

In this case, the probability for Σ to have the metric h_{ij} can be factorized. It is the product of two wave functions Ψ^+ and Ψ^- . These are given by path integrals over all metrics on M^+ and M^- respectively, that induce the given three metric h_{ij} on Σ . In most cases, the two wave functions will be equal and we will drop the superscripts $+$ and $-$. Ψ is called the wave function of the universe. If there are matter fields ϕ , the wave function will also depend on their values ϕ_0 on Σ . But it will not depend explicitly on time because there is no preferred time coordinate in a closed universe. The no boundary proposal implies that the wave function of the universe is given by a path integral over fields on a compact manifold M^+ whose only boundary is the surface Σ . The path integral is taken over all metrics and matter fields on M^+ that agree with the metric h_{ij} and matter fields ϕ_0 on Σ .

One can describe the position of the surface Σ by a function τ of three coordinates x_i on Σ . But the wave function defined by the path integral cannot depend on τ or on the choice of the coordinates x_i . This implies that the wave function Ψ has to obey four functional differential equations. Three of these equations are called the momentum constraint. One can describe the position of the surface Σ by a function τ of three coordinates x_i on Σ . But the wave function defined by the path integral cannot depend on τ or on the choice of the coordinates x_i . This implies that the wave function Ψ has to obey four functional differential equations. Three of these equations are called the *momentum constraint equation*: $\left(\frac{\partial \Psi}{\partial h_{ij}}\right)_{;j} = 0$. They express the fact that the wave function should be the same for different 3 metrics h_{ij} that can be get from each other by transformations of the coordinates x_i . The fourth equation is called the *Wheeler-DeWitt equation*

$$\left(G_{ijkl} \frac{\partial^2}{\partial h_{ij} \partial h_{kl}} - h^{\frac{1}{2}} {}^3R\right) \Psi = 0.$$

It corresponds to the independence of the wave function on τ . One can

think of it as the *Schrödinger equation for the universe*. But there is no time derivative term because the wave function does not depend on time explicitly.

In order to estimate the wave function of the universe, one can use the *saddle point approximation* to the path integral as in the case of black holes. One finds a Euclidean metric g_0 on the manifold M^+ that satisfies the field equations and induces the metric h_{ij} on the boundary Σ . One can then expand the action A in a power series around the background metric g_0 ,

$$A[g] = A[g_0] + \frac{1}{2}\delta g A_2 \delta g + \dots$$

As before, the term linear in the perturbations vanishes. The quadratic term can be regarded as giving the contribution of gravitons on the background and the higher order terms as interactions between the gravitons. These can be ignored when the radius of curvature of the background is large compared to the Planck scale. Therefore, according to [Hawking and Israel (1979); Hawking and Penrose (1996)] we have

$$\Psi \approx \frac{1}{(\det A_2)^{\frac{1}{2}}} \exp(-A[g_0]).$$

Consider now a situation in which there are no matter fields but there is a positive cosmological constant Λ . Let us take the surface Σ to be a three sphere and the metric h_{ij} to be the round three sphere metric of radius a . Then the manifold M^+ bounded by Σ can be taken to be the four ball. The metric that satisfies the field equations is part of a four sphere of radius $\frac{1}{H}$ where $H^2 = \frac{\Lambda}{3}$,

$$A = \frac{1}{16\pi} \int (R - 2\Lambda)(-g)^{\frac{1}{2}} d^4x + \frac{1}{8\pi} \int K(\pm h)^{\frac{1}{2}} d^3x.$$

For a 3-sphere Σ of radius less than $\frac{1}{H}$ there are two possible Euclidean solutions: either M^+ can be less than a hemisphere or it can be more. However there are arguments that show that one should pick the solution corresponding to less than a hemisphere.

One can interpret the wave function Ψ as follows. The real time solution of the Einstein equations with a Λ term and maximal symmetry is de Sitter space (see, e.g., [Witten (1998b)]). This can be embedded as a hyperboloid in five dimensional Minkowski space. Here, we have two choices:

(1) *Lorentzian-de Sitter metric*,

$$ds^2 = -dt^2 + \frac{1}{H^2} \cosh Ht (dr^2 + \sin^2 r (d\theta^2 + \sin^2 \theta d\phi^2)).$$

One can think of it as a closed universe that shrinks down from infinite size to a minimum radius and then expands again exponentially. The metric can be written in the form of a Friedmann universe with scale factor $\cosh Ht$. Putting $\tau = it$ converts the \cosh into \cos giving the Euclidean metric on a four sphere of radius $\frac{1}{H}$.

(2) *Euclidean metric*,

$$ds^2 = d\tau^2 + \frac{1}{H^2} \cos H\tau (dr^2 + \sin^2 r (d\theta^2 + \sin^2 \theta d\phi^2)).$$

Thus one gets the idea that a wave function which varies exponentially with the three metric h_{ij} corresponds to an imaginary time Euclidean metric. On the other hand, a wave function which oscillates rapidly corresponds to a real time Lorentzian metric.

Hawking says: “The Euclidean path integral over all topologically trivial metrics can be done by time slicing and so is unitary when analytically continued to the Lorentzian. On the other hand, the path integral over all topologically non-trivial metrics is asymptotically independent of the initial state. Thus the total path integral is unitary and information is not lost in the formation and evaporation of black holes. The way the information gets out seems to be that a true event horizon never forms, just an apparent horizon.”

Like in the case of the pair creation of black holes, one can describe the spontaneous creation of an exponentially expanding universe. One joins the lower half of the Euclidean four sphere to the upper half of the Lorentzian hyperboloid.

Unlike the black hole pair creation, one couldn't say that the de Sitter universe was created out of field energy in a pre-existing space. Instead, it would quite literally be created out of nothing: not just out of the vacuum but out of absolutely nothing at all because there is nothing outside the universe. In the Euclidean regime, the de Sitter universe is just a closed space like the surface of the Earth but with two more dimensions ([Witten (1998b)]). If the cosmological constant is small compared to the Planck value, the curvature of the Euclidean four sphere should be small. This will mean that the saddle point approximation to the path integral should be

good, and that the calculation of the wave function of the universe will not be affected by our ignorance of what happens in very high curvatures.

One can also solve the field equations for boundary metrics that aren't exactly the round three sphere metric. If the radius of the three sphere is less than $\frac{1}{H}$, the solution is a real Euclidean metric. The action will be real and the wave function will be exponentially damped compared to the round three sphere of the same volume. If the radius of the three sphere is greater than this critical radius there will be two complex conjugate solutions and the wave function will oscillate rapidly with small changes in h_{ij} .

Any measurement made in cosmology can be formulated in terms of the wave function. Thus the no boundary proposal makes cosmology into a science because one can predict the result of any observation. The case we have just been considering of no matter fields and just a cosmological constant does not correspond to the universe we live in. Nevertheless, it is a useful example, both because it is a simple model that can be solved fairly explicitly and because, as we shall see, it seems to correspond to the early stages of the universe.

Although it is not obvious from the wave function, a de Sitter universe has thermal properties rather like a black hole. One can see this by writing the de Sitter metric in a *static form* (rather like the Schwarzschild solution)

$$ds^2 = -(1 - H^2 r^2) dt^2 + (1 - H^2 r^2)^{-1} dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\phi^2).$$

There is an apparent singularity at $r = \frac{1}{H}$. However, as in the Schwarzschild solution, one can remove it by a coordinate transformation and it corresponds to an event horizon.

If one returns to the static form of the de Sitter metric and put $\tau = it$ one gets a Euclidean metric. There is an apparent singularity on the horizon. However, by defining a new radial coordinate and identifying τ with period $\frac{2\pi}{H}$, one gets a regular Euclidean metric which is just the four sphere. Because the imaginary time coordinate is periodic, de Sitter space and all quantum fields in it will behave as if they were at a temperature $\frac{H}{2\pi}$. As we shall see, we can observe the consequences of this temperature in the fluctuations in the microwave background. One can also apply arguments similar to the black hole case to the action of the Euclidean-de Sitter solution [Witten (1998b)]. One finds that it has an intrinsic entropy of $\frac{\pi}{H^2}$, which is a quarter of the area of the event horizon. Again this entropy arises for a topological reason: the Euler number of the four sphere is two. This means that there cannot be a global time coordinate on Euclidean-de Sitter space. One can in-

interpret this *cosmological entropy* as reflecting an observers lack of knowledge of the universe beyond his event horizon [Hawking and Israel (1979); Hawking and Penrose (1996)]:

$$\text{Euclidean metric periodic with period } \frac{2\pi}{H} \Rightarrow \begin{cases} \text{Temperature } T = \frac{H}{2\pi}, \\ \text{Area } \mathcal{A} \text{ of event horizon} = \frac{4\pi}{H^2}, \\ \text{Entropy } S = \frac{\pi}{H^2}. \end{cases}$$

This page intentionally left blank

Chapter 6

Geometrical Path Integrals and Their Applications

The machinery of applied differential geometry, as presented so far, is: (i) rigorous, (ii) elegant, and (iii) powerful – as a tool for understanding, prediction and control of complex nonlinear systems. However, due to its *smooth nature*, it is *limited* to modelling of *deterministic* and *continuous-time* dynamical systems only. Naturally, the question arises: is it possible to extend this smooth machinery, so to be able to effectively deal also with *probabilistic* and *discrete-time* dynamical systems, like e.g., Markov chains? And the answer is: Yes. Namely, in the very core of the XX Century geometrodynamics, there is a powerful *conceptual and computational tool* that is ‘by default’ used as a starting point for virtually every new physical theory – the celebrated *Feynman path integral*. In the path-integral formalism, we first formulate the specific classical *action* of a new theory, and subsequently perform its quantization by means of the associated *amplitude*. This *action–amplitude picture* is the core structure in any new physical theory. Unlike mathematical manifolds, bundles and jets, the path integral is an invention of the physical mind of Richard (Dick) Feynman. Its *virtual paths* are in general neither deterministic nor smooth, although they include bundles and jets of deterministic and smooth paths, as well as Markov chains. Yet, it is essentially an applied differential geometry, with its Riemannian and symplectic versions, among many others. At the beginning, it worked only for *conservative* physical systems. Today it includes also *dissipative structures*, as well as various *sources and sinks*. Its smooth part reveals all celebrated equations of the 20th Century, both classical and quantum. It is the core of modern quantum gravity and string theory. It is arguably the most important construct of mathematical physics. At the edge of a new millennium, if you asked a typical theoretical physicist: what will be your main research tool in the new millennium, he/she would

most probably say: path integral. And today, we see it moving out from physics, into the realm of social sciences. Finally, since Feynman's fairly intuitive invention of the path integral [Feynman (1951)], a lot of research has been done to make it mathematically rigorous (see e.g., [Loo (1999); Loo (2000); Albeverio *et. al.* (1986); Klauder (1997); Klauder (2000); Shabanov and Klauder (1998)]).

6.1 Intuition Behind a Path Integral

6.1.1 Classical Probability Concept

Recall that a *random variable* X is defined by its *distribution function* $f(x)$. Its *probabilistic description* is based on the following rules: (i) $P(X = x_i)$ is the probability that $X = x_i$; and (ii) $P(a \leq X \leq b)$ is the probability that X lies in a closed interval $[a, b]$. Its statistical description is based on: (i) μ_X or $E(X)$ is the mean or expectation of X ; and (ii) σ_X is the standard deviation of X . There are two cases of random variables: discrete and continuous, each having its own probability (and statistics) theory.

6.1.2 Discrete Random Variable

A discrete random variable X has only a countable number of values $\{x_i\}$. Its distribution function $f(x_i)$ has the following properties:

$$P(X = x_i) = f(x_i), \quad f(x_i) \geq 0, \quad \sum_i f(x_i) dx = 1.$$

Statistical description of X is based on its discrete mean value μ_X and standard deviation σ_X , given respectively by

$$\mu_X = E(X) = \sum_i x_i f(x_i), \quad \sigma_X = \sqrt{E(X^2) - \mu_X^2}.$$

6.1.3 Continuous Random Variable

Here $f(x)$ is a piecewise continuous function such that:

$$P(a \leq X \leq b) = \int_a^b f(x) dx, \quad f(x) \geq 0, \quad \int_{-\infty}^{\infty} f(x) dx = \int_{\mathbb{R}} f(x) dx = 1.$$

Statistical description of X is based on its continuous mean μ_X and

standard deviation σ_X , given respectively by

$$\mu_X = E(X) = \int_{-\infty}^{\infty} xf(x) dx, \quad \sigma_X = \sqrt{E(X^2) - \mu_X^2}.$$

Now, let us observe the similarity between the two descriptions. The same kind of similarity between discrete and continuous quantum spectrum stroke Dirac when he suggested the combined integral approach, that he denoted by \oint – meaning ‘both integral and sum at once’: summing over discrete spectrum and integration over continuous spectrum.

To emphasize this similarity even further, as well as to set-up the stage for the path integral, recall the notion of a *cumulative distribution function* of a random variable X , that is a function $F: \mathbb{R} \rightarrow \mathbb{R}$, defined by

$$F(a) = P(X) \leq a.$$

In particular, suppose that $f(x)$ is the distribution function of X . Then

$$F(x) = \sum_{x_i \leq x} f(x_i), \quad \text{or} \quad F(x) = \int_{-\infty}^{\infty} f(t) dt,$$

according to as x is a discrete or continuous random variable. In either case, $F(a) \leq F(b)$ whenever $a \leq b$. Also,

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{and} \quad \lim_{x \rightarrow \infty} F(x) = 1,$$

that is, $F(x)$ is monotonic and its limit to the left is 0 and the limit to the right is 1. Furthermore, its cumulative probability is given by

$$P(a \leq X \leq b) = F(b) - F(a),$$

and the Fundamental Theorem of Calculus tells us that, in the continuum case,

$$f(x) = \partial_x F(x).$$

6.1.4 General Markov Stochastic Dynamics

Recall that *Markov stochastic process* is a random process characterized by a *lack of memory*, i.e., the statistical properties of the immediate future are uniquely determined by the present, regardless of the past [Gardiner (1985)].

For example, a *random walk* is an example of the *Markov chain*, i.e., a discrete-time Markov process, such that the motion of the system in

consideration is viewed as a sequence of states, in which the transition from one state to another depends only on the preceding one, or the probability of the system being in state k depends only on the previous state $k - 1$. The property of a Markov chain of prime importance in biodynamics is the existence of an *invariant distribution of states*: we start with an initial state x_0 whose absolute probability is 1. Ultimately the states should be distributed according to a specified distribution.

Between the pure deterministic dynamics, in which all DOF of the system in consideration are explicitly taken into account, leading to classical dynamical equations, for example in Hamiltonian form (3.34), i.e.,

$$\dot{q}^i = \partial_{p_i} H, \quad \dot{p}_i = -\partial_{q^i} H$$

– and pure stochastic dynamics (Markov process), there is so-called *hybrid dynamics*, particularly *Brownian dynamics*, in which some of DOF are represented only through their *stochastic influence* on others. As an example, suppose a system of particles interacts with a viscous medium. Instead of specifying a detailed interaction of each particle with the particles of the viscous medium, we represent the medium as a *stochastic force* acting on the particle. The stochastic force *reduces the dimensionality* of the dynamics.

Recall that the Brownian dynamics represents the phase-space trajectories of a collection of particles that individually obey *Langevin rate equations* in the field of force (i.e., the particles interact with each other via some deterministic force). For a free particle, the Langevin equation reads [Gardiner (1985)]:

$$m\dot{v} = R(t) - \beta v,$$

where m denotes the mass of the particle and v its velocity. The right-hand side represent the coupling to a *heat bath*; the effect of the random force $R(t)$ is to heat the particle. To balance overheating (on the average), the particle is subjected to *friction* β . In humanoid dynamics this is performed with the Rayleigh–Van der Pol’s *dissipation*. Formally, the solution to the Langevin equation can be written as

$$v(t) = v(0) \exp\left(-\frac{\beta}{m}t\right) + \frac{1}{m} \int_0^t \exp[-(t-\tau)\beta/m] R(\tau) d\tau,$$

where the integral on the right-hand side is a *stochastic integral* and the solution $v(t)$ is a random variable. The stochastic properties of the solution depend significantly on the stochastic properties of the random force

$R(t)$. In the Brownian dynamics the random force $R(t)$ is Gaussian distributed. Then the problem boils down to finding the solution to the Langevin stochastic differential equation with the supplementary condition (mean zero and variance)

$$\langle R(t) \rangle = 0, \quad \langle R(t) R(0) \rangle = 2\beta k_B T \delta(t),$$

where $\langle \cdot \rangle$ denotes the mean value, T is temperature, k_B —*equipartition* (i.e., uniform distribution of energy) coefficient, Dirac $\delta(t)$ —function.

Algorithm for computer simulation of the Brownian dynamics (for a single particle) can be written as [Heermann (1990)]:

- (1) Assign an initial position and velocity.
- (2) Draw a random number from a Gaussian distribution with mean zero and variance.
- (3) Integrate the velocity to get v^{n+1} .
- (4) Add the random component to the velocity.

Another approach to taking account the coupling of the system to a heat bath is to subject the particles to collisions with *virtual particles* [Heermann (1990)]. Such collisions are imagined to affect only momenta of the particles, hence they affect the kinetic energy and introduce fluctuations in the total energy. Each stochastic collision is assumed to be an instantaneous event affecting only one particle.

The collision–coupling idea is incorporated into the Hamiltonian model of dynamics (3.34) by adding a stochastic force $R_i = R_i(t)$ to the \dot{p} equation

$$\dot{q}^i = \partial_{p_i} H, \quad \dot{p}_i = -\partial_{q^i} H + R_i(t).$$

On the other hand, the so-called *Ito stochastic integral* represents a kind of classical Riemann–Stieltjes integral from linear functional analysis, which is (in 1D case) for an arbitrary time–function $G(t)$ defined as the *mean square limit*

$$\int_{t_0}^t G(t) dW(t) = m.s. \lim_{n \rightarrow \infty} \left\{ \sum_{i=1}^n G(t_{i-1}) [W(t_i) - W(t_{i-1})] \right\}.$$

Now, the general ND Markov process can be defined by *Ito* stochastic differential equation (SDE),

$$\begin{aligned} dx_i(t) &= A_i[x^i(t), t] dt + B_{ij}[x^i(t), t] dW^j(t), \\ x^i(0) &= x_{i0}, \quad (i, j = 1, \dots, N) \end{aligned}$$

or corresponding *Ito stochastic integral equation*

$$x^i(t) = x^i(0) + \int_0^t ds A_i[x^i(s), s] + \int_0^t dW^j(s) B_{ij}[x^i(s), s],$$

in which $x^i(t)$ is the variable of interest, the vector $A_i[x(t), t]$ denotes deterministic *drift*, the matrix $B_{ij}[x(t), t]$ represents continuous stochastic *diffusion fluctuations*, and $W^j(t)$ is an N -variable *Wiener process* (i.e., generalized Brownian motion) [Wiener (1961)], and $dW^j(t) = W^j(t+dt) - W^j(t)$.

Now, there are three well-known special cases of the *Chapman-Kolmogorov equation* (see [Gardiner (1985)]):

- (1) When both $B_{ij}[x(t), t]$ and $W(t)$ are zero, i.e., in the case of pure deterministic motion, it reduces to the *Liouville equation*

$$\partial_t P(x', t' | x'', t'') = - \sum_i \frac{\partial}{\partial x^i} \{A_i[x(t), t] P(x', t' | x'', t'')\}.$$

- (2) When only $W(t)$ is zero, it reduces to the *Fokker-Planck equation*

$$\begin{aligned} \partial_t P(x', t' | x'', t'') &= - \sum_i \frac{\partial}{\partial x^i} \{A_i[x(t), t] P(x', t' | x'', t'')\} \\ &+ \frac{1}{2} \sum_{ij} \frac{\partial^2}{\partial x^i \partial x^j} \{B_{ij}[x(t), t] P(x', t' | x'', t'')\}. \end{aligned}$$

- (3) When both $A_i[x(t), t]$ and $B_{ij}[x(t), t]$ are zero, i.e., the state-space consists of integers only, it reduces to the *Master equation* of discontinuous jumps

$$\begin{aligned} \partial_t P(x', t' | x'', t'') &= \\ \int dx \{W(x' | x'', t) P(x', t' | x'', t'') - W(x'' | x', t) P(x', t' | x'', t'')\}. \end{aligned}$$

The *Markov assumption* can now be formulated in terms of the conditional probabilities $P(x^i, t_i)$: if the times t_i increase from right to left, the conditional probability is determined entirely by the knowledge of the most recent condition. Markov process is generated by a set of conditional probabilities whose probability-density $P = P(x', t' | x'', t'')$ evolution obeys

the general *Chapman–Kolmogorov integro–differential equation*

$$\begin{aligned} \partial_t P = & - \sum_i \frac{\partial}{\partial x^i} \{A_i[x(t), t] P\} \\ & + \frac{1}{2} \sum_{ij} \frac{\partial^2}{\partial x^i \partial x^j} \{B_{ij}[x(t), t] P\} + \int dx \{W(x'|x'', t) P - W(x''|x', t) P\} \end{aligned}$$

including *deterministic drift*, *diffusion fluctuations* and *discontinuous jumps* (given respectively in the first, second and third terms on the r.h.s.).

It is this general Chapman–Kolmogorov integro–differential equation, with its conditional probability density evolution, $P = P(x', t'|x'', t'')$, that we are going to model by various forms of the Feynman path integral, providing us with the physical insight behind the abstract (conditional) probability densities.

6.1.5 Quantum Probability Concept

An alternative concept of probability, the so-called *quantum probability*, is based on the following physical facts (elaborated in detail in this section):

- (1) The *time-dependent Schrödinger equation* represents a *complex-valued generalization* of the real-valued *Fokker–Planck equation* for describing the spatio-temporal *probability density function* for the system exhibiting *continuous-time Markov stochastic process*.
- (2) The *Feynman path integral* \oint is a generalization of the time-dependent Schrödinger equation, including both continuous-time and discrete-time Markov stochastic processes.
- (3) Both Schrödinger equation and path integral give ‘physical description’ of any system they are modelling in terms of its physical energy, instead of an abstract probabilistic description of the Fokker–Planck equation.

Therefore, the *Feynman path integral* \oint , as a generalization of the time-dependent Schrödinger equation, gives a unique physical description for the general Markov stochastic process, in terms of the physically based generalized probability density functions, valid both for continuous-time and discrete-time Markov systems.

Basic consequence: a different way for calculating probabilities. The difference is rooted in the fact that *sum of squares is different from the square of sums*, as is explained in the following text.

Namely, in Dirac–Feynman quantum formalism, each possible route from the initial system state A to the final system state B is called a *history*. This history comprises any kind of a route (see Figure 6.1), ranging from continuous and smooth deterministic (mechanical–like) paths to completely discontinues and random Markov chains (see, e.g., [Gardiner (1985)]). Each history (labelled by index i) is quantitatively described by a *complex number*¹ z_i called the ‘individual transition amplitude’. Its absolute square, $|z_i|^2$, is called the *individual transition probability*. Now, the *total transition amplitude* is the sum of all individual transition amplitudes, $\sum_i z_i$, called the *sum-over-histories*. The absolute square of this sum-over-histories, $|\sum_i z_i|^2$, is the *total transition probability*.

In this way, the overall probability of the system’s transition from some initial state A to some final state B is given *not* by adding up the probabilities for each history–route, but by ‘head-to-tail’ adding up the sequence of amplitudes making-up each route first (i.e., performing the sum-over-histories) – to get the total amplitude as a ‘resultant vector’, and then squaring the total amplitude to get the overall transition probability.

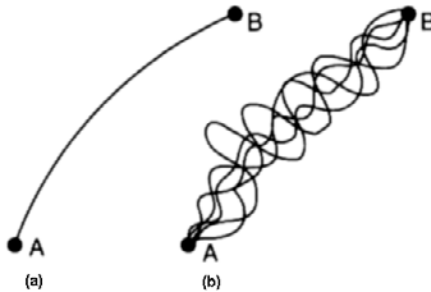


Fig. 6.1 Two ways of physical *transition* from an *initial state* A to the corresponding *final state* B . (a) Classical physics proposes a *single deterministic trajectory*, minimizing the total system’s energy. (b) Quantum physics proposes a *family of Markov stochastic histories*, namely *all possible routes* from A to B , both continuous–time and discrete–time Markov chains, each giving an equal contribution to the total *transition probability*.

¹Recall that a *complex number* $z = x + iy$, where $i = \sqrt{-1}$ is the *imaginary unit*, x is the *real part* and y is the *imaginary part*, can be represented also in its *polar form*, $z = r(\cos \theta + i \sin \theta)$, where the radius vector in the complex plane, $r = |z| = \sqrt{x^2 + y^2}$, is the *modulus* or *amplitude*, and angle θ is the *phase*; as well as in its exponential form $z = re^{i\theta}$. In this way, complex numbers actually represent 2D vectors with usual vector ‘head-to-tail’ addition rule.

6.1.6 Quantum Coherent States

Recall that a *quantum coherent state* is a specific kind of quantum state of the quantum harmonic oscillator whose dynamics most closely resemble the oscillating behavior of a classical harmonic oscillator. It was the first example of quantum dynamics when Erwin Schrödinger derived it in 1926 while searching for solutions of the *Schrödinger equation* that satisfy the *correspondence principle*. The quantum harmonic oscillator and hence, the coherent state, arise in the quantum theory of a wide range of physical systems. For instance, a coherent state describes the oscillating motion of the particle in a quadratic potential well. In the quantum electrodynamics and other bosonic quantum field theories they were introduced by the 2005 Nobel Prize winning work of Roy Glauber in 1963 [Glauber (1963a); Glauber (1963b)]. Here the coherent state of a field describes an oscillating field, the closest quantum state to a classical sinusoidal wave such as a continuous laser wave.

In classical optics, light is thought of as electromagnetic waves radiating from a source. Specifically, coherent light is thought of as light that is emitted by many such sources that are in phase. For instance, a light bulb radiates light that is the result of waves being emitted at all the points along the filament. Such light is incoherent because the process is highly random in space and time. On the other hand, in a laser, light is emitted by a carefully controlled system in processes that are not random but interconnected by stimulation and the resulting light is highly ordered, or coherent. Therefore a coherent state corresponds closely to the quantum state of light emitted by an ideal laser. Semi-classically we describe such a state by an electric field oscillating as a stable wave. Contrary to the coherent state, which is the most wave-like quantum state, the *Fock state* (e.g., a single photon) is the most particle-like state. It is indivisible and contains only one quanta of energy. These two states are examples of the opposite extremes in the concept of *wave-particle duality*. A coherent state distributes its quantum-mechanical uncertainty equally, which means that the phase and amplitude uncertainty are approximately equal. Conversely, in a single-particle state the phase is completely uncertain.

Formally, the coherent state $|\alpha\rangle$ is defined to be the eigenstate of the annihilation operator a , i.e., $a|\alpha\rangle = \alpha|\alpha\rangle$. Note that since a is not Hermitian, $\alpha = |\alpha|e^{i\theta}$ is complex. $|\alpha|$ and θ are called the *amplitude* and *phase* of the state.

Physically, $a|\alpha\rangle = \alpha|\alpha\rangle$ means that a coherent state is left unchanged

by the detection (or annihilation) of a particle. Consequently, in a coherent state, one has exactly the same probability to detect a second particle. Note, this condition is necessary for the coherent state's *Poisson detection statistics*. Compare this to a single-particle's Fock state: Once one particle is detected, we have zero probability of detecting another.

Now, recall that a *Bose–Einstein condensate* (BEC) is a collection of boson atoms that are all in the same quantum state. An approximate theoretical description of its properties can be derived by assuming the BEC is in a coherent state. However, unlike photons, atoms interact with each other so it now appears that it is more likely to be one of the *squeezed coherent states* (see [Breitenbach *et al.* (1997)]). In quantum field theory and string theory, a generalization of coherent states to the case of infinitely many degrees of freedom is used to define a *vacuum state* with a different vacuum expectation value from the original vacuum.

6.1.7 Dirac's $\langle \text{bra} | \text{ket} \rangle$ Transition Amplitude

Now, we are ready to move-on into the realm of quantum mechanics. Recall that [Dirac (1982)] described behavior of quantum systems in terms of complex-valued *ket-vectors* $|A\rangle$ living in the Hilbert space \mathcal{H} , and their duals, *bra-covectors* $\langle B|$ (i.e., 1-forms) living in the *dual* Hilbert space \mathcal{H}^* .² The *Hermitian inner product* of kets and bras, the *bra–ket* $\langle B|A\rangle$, is a *complex number*, which is the evaluation of the ket $|A\rangle$ by the bra $\langle B|$. This complex number, say $re^{i\theta}$ represents the system's *transition*

²Recall that a *norm* on a complex vector space \mathcal{H} is a mapping from \mathcal{H} into the complex numbers, $\|\cdot\| : \mathcal{H} \rightarrow \mathbb{C}$; $h \mapsto \|h\|$, such that the following set of norm-axioms hold:

- (N1) $\|h\| \geq 0$ for all $h \in \mathcal{H}$ and $\|h\| = 0$ implies $h = 0$ (positive definiteness);
- (N2) $\|\lambda h\| = |\lambda| \|h\|$ for all $h \in \mathcal{H}$ and $\lambda \in \mathbb{C}$ (homogeneity); and
- (N3) $\|h_1 + h_2\| \leq \|h_1\| + \|h_2\|$ for all $h_1, h_2 \in \mathcal{H}$ (triangle inequality). The pair $(\mathcal{H}, \|\cdot\|)$ is called a *normed space*.

A *Hermitian inner product* on a complex vector space \mathcal{H} is a mapping $\langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{C}$ such that the following set of inner-product-axioms hold:

- (IP1) $\langle h h_1 + h_2 \rangle = \langle h h_1 \rangle + \langle h h_2 \rangle$;
- (IP2) $\langle \alpha h, h_1 \rangle = \alpha \langle h, h_1 \rangle$;
- (IP3) $\langle h_1, h_2 \rangle = \overline{\langle h_2, h_1 \rangle}$ (so $\langle h, h \rangle$ is real);
- (IP4) $\langle h, h \rangle \geq 0$ and $\langle h, h \rangle = 0$ provided $h = 0$.

The *standard inner product* on the product space $\mathbb{C}^n = \mathbb{C} \times \cdots \times \mathbb{C}$ is defined by $\langle z, w \rangle = \sum_{i=1}^n z_i \overline{w_i}$, and axioms (IP1)–(IP4) are readily checked. Also \mathbb{C}^n is a normed space with $\|z\|^2 = \sum_{i=1}^n |z_i|^2$. The pair $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ is called an *inner product space*.

Let $(\mathcal{H}, \|\cdot\|)$ be a normed space. If the corresponding metric d is complete, we say $(\mathcal{H}, \|\cdot\|)$ is a Banach space. If $(\mathcal{H}, \|\cdot\|)$ is an inner product space whose corresponding metric is complete, we say $(\mathcal{H}, \|\cdot\|)$ is a *Hilbert space*.

amplitude³ from its *initial state* A to its *final state* B^4 , i.e.,

$$\text{Transition Amplitude} = \langle B|A \rangle = re^{i\theta}.$$

That is, there is a process that can mediate a transition of a system from initial state A to the final state B and the amplitude for this transition equals $\langle B|A \rangle = re^{i\theta}$. The absolute square of the amplitude, $|\langle B|A \rangle|^2$ represents the *transition probability*. Therefore, the probability of a transition event equals the absolute square of a complex number, i.e.,

$$\text{Transition Probability} = |\langle B|A \rangle|^2 = |re^{i\theta}|^2.$$

These complex amplitudes obey the usual *laws of probability*: when a transition event can happen in alternative ways then we add the complex numbers,

$$\langle B_1|A_1 \rangle + \langle B_2|A_2 \rangle = r_1e^{i\theta_1} + r_2e^{i\theta_2},$$

and when it can happen only as a succession of intermediate steps then we multiply the complex numbers,

$$\langle B|A \rangle = \langle B|c \rangle \langle c|A \rangle = (r_1e^{i\theta_1})(r_2e^{i\theta_2}) = r_1r_2e^{i(\theta_1+\theta_2)}.$$

In general,

- (1) The amplitude for n *mutually alternative processes* equals the *sum* $\sum_{k=1}^n r_k e^{i\theta_k}$ of the amplitudes for the alternatives; and
- (2) If transition from A to B occurs in a *sequence of m steps*, then the total transition amplitude equals the *product* $\prod_{j=1}^m r_j e^{i\theta_j}$ of the amplitudes of the steps.

Formally, we have the so-called *expansion principle*, including both

³Transition amplitude is otherwise called *probability amplitude*, or just *amplitude*.

⁴Recall that in quantum mechanics, complex numbers are regarded as the *vacuum-state*, or the *ground-state*, and the entire amplitude $\langle b|a \rangle$ is a *vacuum-to-vacuum amplitude* for a process that includes the *creation* of the state a , its *transition* to b , and the *annihilation* of b to the vacuum once more.

products and sums,⁵

$$\langle B|A \rangle = \sum_{i=1}^n \langle B|c_i \rangle \langle c_i|A \rangle. \quad (6.1)$$

6.1.8 Feynman's Sum-over-Histories

Now, *iterating* the Dirac's expansion principle (6.1) over a *complete set of all possible states* of the system, leads to the simplest form of the *Feynman path integral*, or, *sum-over-histories*. Imagine that the *initial* and *final* states, A and B , are points on the vertical lines $x = 0$ and $x = n + 1$, respectively, in the $x - y$ plane, and that $(c(k)_{i(k)}, k)$ is a given point on the line $x = k$ for $0 < i(k) < m$ (see Figure 6.2). Suppose that the sum of projectors for each intermediate state is complete⁶ Applying the completeness iteratively, we get the following expression for the transition amplitude:

$$\langle B|A \rangle = \sum \sum \dots \sum \langle B|c(1)_{i(1)} \rangle \langle c(1)_{i(1)}|c(2)_{i(2)} \rangle \dots \langle c(n)_{i(n)}|A \rangle,$$

where the sum is taken over all $i(k)$ ranging between 1 and m , and k ranging between 1 and n . Each term in this sum can be construed as a *combinatorial route* from A to B in the 2D space of the $x - y$ plane. Thus the transition amplitude for the system going from some initial state A to some final state B is seen as a summation of contributions from *all the routes* connecting A to B .

Feynman used this description to produce his celebrated *path integral* expression for a transition amplitude (see, e.g., [Grosche and Steiner (1998)]). His path integral takes the form

$$\text{Transition Amplitude} = \langle B|A \rangle = \oint \mathcal{D}[x] e^{iS[x]}, \quad (6.2)$$

⁵In Dirac's language, the *completeness* of intermediate states becomes the statement that a certain sum of projectors is equal to the identity. Namely, suppose that $\sum_i |c_i \rangle \langle c_i| = 1$ with $\langle c_i|c_i \rangle = 1$ for each i . Then

$$\langle b|a \rangle = \langle b||a \rangle = \langle b| \sum_i |c_i \rangle \langle c_i| |a \rangle = \sum_i \langle b|c_i \rangle \langle c_i|a \rangle.$$

⁶We assume that following sum is equal to one, for each k from 1 to $n - 1$:

$$|c(k)_1 \rangle \langle c(k)_1| + \dots + |c(k)_m \rangle \langle c(k)_m| = 1.$$

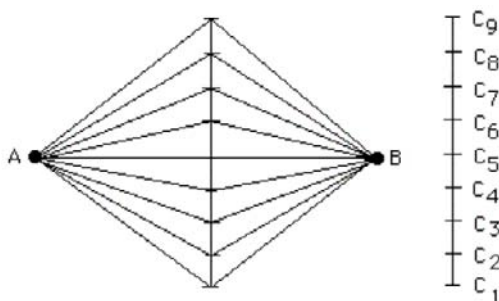


Fig. 6.2 Analysis of all possible routes from the source A to the detector B is simplified to include only double straight lines (in a plane).

where the sum–integral \int is taken over all possible routes $x = x(t)$ from the initial point $A = A(t_{ini})$ to the final point $B = B(t_{fin})$, and $\mathcal{S} = \mathcal{S}[x]$ is the classical *action* for a particle to travel from A to B along a given extremal path x . In this way, Feynman took seriously Dirac’s conjecture interpreting the exponential of the classical *action functional* ($\mathcal{D}e^{i\mathcal{S}}$), resembling a complex number ($re^{i\theta}$), as an *elementary amplitude*. By integrating this elementary amplitude, $\mathcal{D}e^{i\mathcal{S}}$, over the infinitude of all possible histories, we get the total system’s transition amplitude.⁷

⁷For the quantum physics associated with a classical (Newtonian) particle the action \mathcal{S} is given by the integral along the given route from a to b of the difference $T - V$ where T is the classical kinetic energy and V is the classical potential energy of the particle. The beauty of Feynman’s approach to quantum physics is that it shows the relationship between the classical and the quantum in a particularly transparent manner. Classical motion corresponds to those regions where all nearby routes contribute constructively to the summation. This classical path occurs when the *variation of the action* is null. To ask for those paths where the variation of the action is zero is a problem in the calculus of variations, and it leads directly to Newton’s equations of motion (derived using the Euler–Lagrangian equations). Thus with the appropriate choice of action, classical and quantum points of view are unified.

Also, a discretization of the Schrodinger equation

$$i\hbar \frac{d\psi}{dt} = -\frac{\hbar^2}{2m} \frac{d^2\psi}{dx^2} + V\psi,$$

leads to a sum–over–histories that has a discrete path integral as its solution. Therefore, the transition amplitude is equivalent to the wave ψ . The particle travelling on the x –axis is executing a one–step *random walk*, see Figure 6.3.

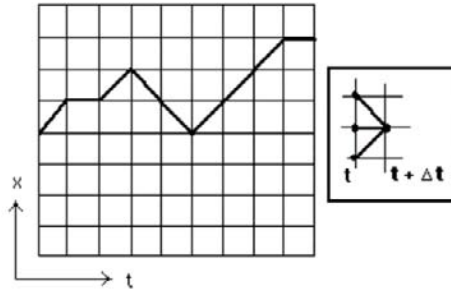


Fig. 6.3 Random walk (a particular case of Markov chain) on the x -axis.

6.1.9 The Basic Form of a Path Integral

In Feynman's version of non-relativistic quantum mechanics, the time evolution $\psi(x', t') \mapsto \psi(x'', t'')$ of the wave function $\psi = \psi(x, t)$ of the elementary 1D particle may be described by the integral equation [Grosche and Steiner (1998)]

$$\psi(x'', t'') = \int_{\mathbb{R}} K(x'', x'; t'', t') \psi(x', t'), \quad (6.3)$$

where the *propagator* or *Feynman kernel* $K = K(x'', x'; t'', t')$ is defined through a limiting procedure,

$$K(x'', x'; t'', t') = \lim_{\epsilon \rightarrow 0} A^{-N} \prod_{k=1}^{N-1} \int dx_k e^{i \sum_{j=0}^{N-1} \epsilon L(x_{j+1}, (x_{j+1} - x_j)/\epsilon)}. \quad (6.4)$$

The time interval $t'' - t'$ has been discretized into N steps of length $\epsilon = (t'' - t')/N$, and the r.h.s. of (6.4) represents an integral over all piecewise linear paths $x(t)$ of a 'virtual' particle propagating from x' to x'' , illustrated in Figure 6.4.

The prefactor A^{-N} is a normalization and L denotes the Lagrangian function of the particle. Knowing the propagator G is tantamount to having solved the quantum dynamics. This is the simplest instance of a path integral, and is often written schematically as

$$K(x', t'; x'', t'') = \int \mathcal{D}[x(t)] e^{iS[x(t)]},$$

where $\mathcal{D}[x(t)]$ is a functional measure on the 'space of all paths', and the exponential weight depends on the classical action $S[x(t)]$ of a path. Recall also that this procedure can be defined in a mathematically clean way if

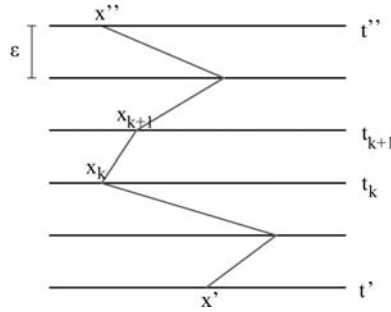


Fig. 6.4 A piecewise linear particle path contributing to the discrete Feynman propagator.

we Wick-rotate the time variable t to imaginary values $t \mapsto \tau = it$, thereby making all integrals real [Reed and Simon (1975)].

6.1.10 Application: *Adaptive Path Integral*

Now, we can extend the Feynman sum-over-histories (6.2), by adding the synaptic-like weights $w^i = w^i(t)$ into the measure $\mathcal{D}[x]$, to get the *adaptive path integral*:

$$\text{Adaptive Transition Amplitude} = \langle B|A \rangle_w = \oint \mathcal{D}[w, x] e^{iS[x]}, \quad (6.5)$$

where the *adaptive measure* $\mathcal{D}[w, x]$ is defined by the weighted product (of discrete time steps)

$$\mathcal{D}[w, x] = \lim_{n \rightarrow \infty} \prod_{t=1}^n w^i(t) dx^i(t). \quad (6.6)$$

In (6.6) the *synaptic weights* $w^i = w^i(t)$ are updated by the unsupervised *Hebbian-like learning* rule [Hebb (1949)]:

$$w^i(t+1) = w^i(t) + \frac{\sigma}{\eta} (w_d^i(t) - w_a^i(t)), \quad (6.7)$$

where $\sigma = \sigma(t)$, $\eta = \eta(t)$ represent local *signal* and *noise* amplitudes, respectively, while superscripts d and a denote *desired* and *achieved* system states, respectively. Theoretically, equations (6.5–6.7) define an ∞ -dimensional complex-valued neural network.⁸ Practically, in a com-

⁸For details on complex-valued neural networks, see e.g., complex-domain extension of the standard backpropagation learning algorithm [Georgiou and Koutsougeras (1992); Benvenuto and Piazza (1992)].

puter simulation we can use $10^7 \leq n \leq 10^8$, approaching the number of neurons in the brain. Such equations are usually solved using Markov–Chain Monte–Carlo methods on parallel (cluster) computers (see, e.g., [Wehner and Wolfer (1983a); Wehner and Wolfer (1983b)]).

6.2 Path Integral History

6.2.1 Extract from Feynman’s Nobel Lecture

In his Nobel Lecture, December 11, 1965, Richard (Dick) Feynman said that he and his PhD supervisor, John Wheeler, had found the *action* $A = A[x; t_i, t_j]$, directly involving the *motions of the charges only*,⁹

$$A[x; t_i, t_j] = m_i \int (\dot{x}_\mu^i \dot{x}_\mu^i)^{\frac{1}{2}} dt_i + \frac{1}{2} e_i e_j \int \int \delta(I_{ij}^2) \dot{x}_\mu^i(t_i) \dot{x}_\mu^j(t_j) dt_i dt_j$$

with $(i \neq j)$ (6.8)

$$I_{ij}^2 = [x_\mu^i(t_i) - x_\mu^j(t_j)] [x_\mu^i(t_i) - x_\mu^j(t_j)],$$

where $x_\mu^i = x_\mu^i(t_i)$ is the four–vector *position* of the i th particle as a function of the proper time t_i , while $\dot{x}_\mu^i(t_i) = dx_\mu^i(t_i)/dt_i$ is the *velocity* four–vector.

The first term in the action $A[x; t_i, t_j]$ (6.8) is the integral of the proper time t_i , the *ordinary action of relativistic mechanics of free particles of mass m_i* (summation over μ). The second term in the action $A[x; t_i, t_j]$ (6.8) represents the *electrical interaction of the charges*. It is summed over each pair of charges (the factor $\frac{1}{2}$ is to count each pair once, the term $i = j$ is omitted to avoid self–action). The *interaction is a double integral over a delta function of the square of space–time interval I^2 between two points on the paths*. Thus, interaction occurs only when this interval vanishes, that is, *along light cones* (see [Wheeler and Feynman (1949)]).

Feynman comments here: “The fact that the interaction is exactly one–half advanced and half–retarded meant that we could write such a principle of least action, whereas interaction via retarded waves alone cannot be written in such a way. So, all of classical electrodynamics was contained in this very simple form.”

“...The problem is only to make a quantum theory, which has as its classical analog, this expression (6.8). Now, there is no unique way to make a quantum theory from classical mechanics, although all the textbooks make

⁹ *Wheeler–Feynman Idea* [Wheeler and Feynman (1949)] “The energy tensor can be regarded only as a provisional means of representing matter. In reality, *matter consists of electrically charged particles*.”

believe there is. What they would tell you to do, was find the momentum variables and replace them by $(\hbar/i)(\partial/\partial x)$, but I couldn't find a momentum variable, as there wasn't any."

"The character of quantum mechanics of the day was to write things in the famous *Hamiltonian way* (in the form of Schrödinger equation), which described how the wave function changes from instant to instant, and in terms of the Hamiltonian operator H . If the classical physics could be reduced to a Hamiltonian form, everything was all right. Now, least action does not imply a Hamiltonian form if the action is a function of anything more than positions and velocities at the same moment. If the action is of the form of the integral of the Lagrangian $L = L(\dot{x}, x)$, a function of the velocities and positions at the same time t ,

$$S[x] = \int L(\dot{x}, x) dt, \quad (6.9)$$

then you can start with the Lagrangian L and then create a Hamiltonian H and work out the quantum mechanics, more or less uniquely. But the action $A[x; t_i, t_j]$ (6.8) involves the key variables, positions (and velocities), at two different times t_i and t_j and therefore, it was not obvious what to do to make the quantum-mechanical analogue..."

So, Feynman was looking for the action integral in quantum mechanics. He says: "...I simply turned to Professor Jehle and said, 'Listen, do you know any way of doing quantum mechanics, starting with action – where the action integral comes into the quantum mechanics?' 'No", he said, 'but Dirac has a paper in which the Lagrangian, at least, comes into quantum mechanics." What Dirac said was the following: There is in quantum mechanics a very important quantity which carries the wave function from one time to another, besides the differential equation but equivalent to it, a kind of a kernel, which we might call $K(x', x)$, which carries the wave function $\psi(x)$ known at time t , to the wave function $\psi(x')$ at time $t + \varepsilon$,

$$\psi(x', t + \varepsilon) = \int K(x', x) \psi(x, t) dx.$$

Dirac points out that this function K was analogous to the quantity in classical mechanics that you would calculate if you took the exponential of [$i\varepsilon$ multiplied by the Lagrangian $L(\dot{x}, x)$], imagining that these two positions x, x' corresponded to t and $t + \varepsilon$. In other words,

$$K(x', x) \quad \text{is analogous to} \quad e^{i\varepsilon L(\frac{x' - x}{\varepsilon}, x)/\hbar}.$$

So, Feynman continues: “What does he mean, they are analogous; what does that mean, *analogous*? What is the use of that?” Professor Jehle said, ‘You Americans! You always want to find a use for everything!’ I said that I thought that Dirac must mean that they were *equal*. ‘No’, he explained, ‘he doesn’t mean they are equal.’ ‘Well’, I said, ‘Let’s see what happens if we make them equal.’

“So, I simply put them equal, taking the simplest example where the Lagrangian is

$$L = \frac{1}{2}M\dot{x}^2 - V(x),$$

but soon found I had to put a constant of proportionality N in, suitably adjusted. When I substituted for K to get

$$\psi(x', t + \varepsilon) = \int N \exp \left[\frac{i\varepsilon}{\hbar} L\left(\frac{x' - x}{\varepsilon}, x\right) \right] \psi(x, t) dx \quad (6.10)$$

and just calculated things out by Taylor series expansion, *out came the Schrödinger equation*. So, I turned to Professor Jehle, not really understanding, and said, ‘Well, you see, Dirac meant that they were proportional.’ Professor Jehle’s eyes were bugging out – he had taken out a little notebook and was rapidly copying it down from the blackboard, and said, ‘No, no, this is an important discovery. You Americans are always trying to find out how something can be used. That’s a good way to discover things!’ So, I thought I was finding out what Dirac meant, but, as a matter of fact, had made the discovery that what Dirac thought was analogous, was, in fact, equal. I had then, at least, the connection between the Lagrangian and quantum mechanics, but still with wave functions and infinitesimal times.”

“It must have been a day or so later when I was lying in bed thinking about these things, that I imagined what would happen if I wanted to calculate the wave function at a finite interval later. I would put one of these factors $e^{i\varepsilon L}$ in here, and that would give us the wave functions the next moment, $t + \varepsilon$, and then I could substitute that back into (6.10) to get another factor of $e^{i\varepsilon L}$ and give us the wave function the next moment, $t + 2\varepsilon$, and so on and so on. In that way I found myself thinking of a large number of integrals, one after the other in sequence. In the integrand was the product of the exponentials, which was the exponential of the sum of terms like εL . Now, L is the Lagrangian and ε is like the time interval dt , so that if you took a sum of such terms, that’s exactly like an integral. That’s like Riemann’s formula for the integral $\int L dt$, you just take the value at

each point and add them together. We are to take the limit as $\varepsilon \rightarrow 0$. Therefore, the connection between the wave function of one instant and the wave function of another instant a finite time later could be get by an infinite number of integrals (because ε goes to zero), of exponential where S is the action expression (6.9). At last, I had succeeded in representing quantum mechanics directly in terms of the action $S[x]$."

Fully satisfied, Feynman comments: "This led later on to the idea of the *transition amplitude* for a path: that for each possible way that the particle can go from one point to another in space-time, there's an amplitude. That amplitude is \mathbf{e} to the power of [i/\hbar times the action $S[x]$ for the path], i.e., $\mathbf{e}^{iS[x]/\hbar}$. Amplitudes from various paths superpose by addition. This then is another, a third way, of describing quantum mechanics, which looks quite different from that of Schrödinger or Heisenberg, but which is equivalent to them."

"...Now immediately after making a few checks on this thing, what we wanted to do, was to substitute the action $A[x; t_i, t_j]$ (6.8) for the other $S[x]$ (6.9). The first trouble was that I could not get the thing to work with the relativistic case of spin one-half. However, although I could deal with the matter only non-relativistically, I could deal with the light or the photon interactions perfectly well by just putting the interaction terms of (6.8) into any action, replacing the mass terms by the non-relativistic $Ldt = \frac{1}{2}M\dot{x}^2dt$,

$$A[x; t_i, t_j] = \frac{1}{2} \sum_i m_i \int (\dot{x}_\mu^i)^2 dt_i + \frac{1}{2} \sum_{i,j(i \neq j)} e_i e_j \int \int \delta(I_{ij}^2) \dot{x}_\mu^i(t_i) \dot{x}_\mu^j(t_j) dt_i dt_j.$$

When the action has a delay, as it now had, and involved more than one time, I had to lose the idea of a wave function. That is, I could no longer describe the program as: given the amplitude for all positions at a certain time to calculate the amplitude at another time. However, that didn't cause very much trouble. It just meant developing a new idea. *Instead of wave functions we could talk about this: that if a source of a certain kind emits a particle, and a detector is there to receive it, we can give the amplitude that the source will emit and the detector receive, $\mathbf{e}^{iA[x;t_i,t_j]/\hbar}$.* We do this without specifying the exact instant that the *source* emits or the exact instant that any *detector* receives, without trying to specify the state of anything at any particular time in between, but by just finding the *amplitude for the complete experiment*. And, then we could discuss how that amplitude would change if you had a scattering sample in between, as

you rotated and changed angles, and so on, without really having any wave functions...It was also possible to discover what the old concepts of energy and momentum would mean with this generalized action. And, so I believed that I had a quantum theory of classical electrodynamics – or rather of this new classical electrodynamics described by the action $A[x; t_i, t_j]$ (6.8)..."

6.2.2 Lagrangian Path Integral

Dirac and Feynman first developed the Lagrangian approach to functional integration. To review this approach, we start with the time-dependent *Schrödinger equation*

$$i\hbar \partial_t \psi(x, t) = -\partial_{x^2} \psi(x, t) + V(x) \psi(x, t)$$

appropriate to a particle of mass m moving in a potential $V(x)$, $x \in \mathbb{R}$. A solution to this equation can be written as an integral (see e.g., [Klauder (1997); Klauder (2000)]),

$$\psi(x'', t'') = \int K(x'', t''; x', t') \psi(x', t') dx',$$

which represents the wave function $\psi(x'', t'')$ at time t'' as a linear superposition over the wave function $\psi(x', t')$ at the initial time t' , $t' < t''$. The integral kernel $K(x'', t''; x', t')$ is known as the *propagator*, and according to Feynman [Feynman (1948)] it may be given by

$$K(x'', t''; x', t') = \mathcal{N} \int \mathcal{D}[x] e^{(i/\hbar) \int [(m/2) \dot{x}^2(t) - V(x(t))] dt},$$

which is a formal expression symbolizing an integral over a suitable set of paths. This integral is supposed to run over all continuous paths $x(t)$, $t' \leq t \leq t''$, where $x(t'') = x''$ and $x(t') = x'$ are fixed end points for all paths. Note that the integrand involves the *classical Lagrangian* for the system.

To overcome the convergence problems, Feynman adopted a *lattice regularization* as a procedure to yield well-defined integrals which was then followed by a limit as the lattice spacing goes to zero called the continuum limit. With $\varepsilon > 0$ denoting the lattice spacing, the details regarding the

lattice regularization procedure are given by

$$K(x'', t''; x', t') = \lim_{\varepsilon \rightarrow 0} (m/2\pi i \hbar \varepsilon)^{(N+1)/2} \int \cdots \\ \cdots \int \exp\left\{ (i/\hbar) \sum_{l=0}^N [(m/2\varepsilon)(x_{l+1} - x_l)^2 - \varepsilon V(x_l)] \right\} \prod_{l=1}^N dx_l ,$$

where $x_{N+1} = x''$, $x_0 = x'$, and $\varepsilon \equiv (t'' - t')/(N + 1)$, $N \in \{1, 2, 3, \dots\}$. In this version, at least, we have an expression that has a reasonable chance of being well defined, provided, that one interprets the conditionally convergent integrals involved in an appropriate manner. One common and fully acceptable interpretation adds a convergence factor to the exponent of the preceding integral in the form $-(\varepsilon^2/2\hbar) \sum_{l=1}^N x_l^2$, which is a term that formally makes no contribution to the final result in the continuum limit save for ensuring that the integrals involved are now rendered absolutely convergent.

6.2.3 Hamiltonian Path Integral

It is necessary to retrace history at this point to recall the introduction of the *phase-space path integral* by Feynman [Feynman (1951); Grosche and Steiner (1998)]. In Appendix B to this article, Feynman introduced a formal expression for the configuration or q -space propagator given by (see e.g., [Klauder (1997); Klauder (2000)])

$$K(q'', t''; q', t') = \mathcal{M} \int \mathcal{D}[p] \mathcal{D}[q] \exp\left\{ (i/\hbar) \int [p \dot{q} - H(p, q)] dt \right\}.$$

In this equation one is instructed to integrate over all paths $q(t)$, $t' \leq t \leq t''$, with $q(t'') \equiv q''$ and $q(t') \equiv q'$ held fixed, as well as to integrate over all paths $p(t)$, $t' \leq t \leq t''$, without restriction.

It is widely appreciated that the phase-space path integral is more generally applicable than the original, Lagrangian, version of the path integral. For example, the original configuration space path integral is satisfactory for Lagrangians of the general form

$$L(x) = \frac{1}{2} m \dot{x}^2 + A(x) \dot{x} - V(x) ,$$

but it is unsuitable, for example, for the case of a relativistic particle with the Lagrangian

$$L(x) = -m \sqrt{c^2 - \dot{x}^2}$$

expressed in units where the speed of light is unity. For such a system – as well as many more general expressions – the phase-space form of the path integral is to be preferred. In particular, for the relativistic free particle, the phase-space path integral

$$\mathcal{M} \int \mathcal{D}[p] \mathcal{D}[q] \exp\{ (i/\hbar) \int [p \dot{q} - q r t p^2 + m^2] dt \},$$

is readily evaluated and induces the correct propagator.

6.2.4 Feynman–Kac Formula

Through his own research, M. Kac was fully aware of *Wiener’s theory of Brownian motion* and the *associated diffusion equation* that describes the corresponding *distribution function*. Therefore, it is not surprising that he was well prepared to give a path integral expression in the sense of Feynman for an equation similar to the time-dependent Schrödinger equation save for a rotation of the time variable by $-\pi/2$ in the complex plane, namely, by the change $t \rightarrow -it$ (see e.g., [Klauder (1997); Klauder (2000)]). In particular, Kac [Kac (1951)] considered the equation

$$\partial_t \rho(x, t) = \partial_{x^2} \rho(x, t) - V(x) \rho(x, t). \quad (6.11)$$

This equation is analogous to Schrödinger equation but differs from it in certain details. Besides certain constants which are different, and the change $t \rightarrow -it$, the nature of the dependent variable function $\rho(x, t)$ is quite different from the normal quantum mechanical wave function. For one thing, if the function ρ is initially real it will remain real as time proceeds. Less obvious is the fact that if $\rho(x, t) \geq 0$ for all x at some time t , then the function will continue to be nonnegative for all time t . Thus we can interpret $\rho(x, t)$ more like a probability density; in fact in the special case that $V(x) = 0$, then $\rho(x, t)$ is the probability density for a Brownian particle which underlies the *Wiener measure*. In this regard, ν is called the diffusion constant.

The fundamental solution of (6.11) with $V(x) = 0$ is readily given as

$$W(x, T; y, 0) = \frac{1}{q r t 2 \pi \nu T} \exp \left(- \frac{(x - y)^2}{2 \nu T} \right),$$

which describes the solution to the diffusion equation subject to the initial

condition

$$\lim_{T \rightarrow 0^+} W(x, T; y, 0) = \delta(x - y) .$$

Moreover, it follows that the solution of the diffusion equation for a general initial condition is given by

$$\rho(x'', t'') = \int W(x'', t''; x', t') \rho(x', t') dx' .$$

Iteration of this equation N times, with $\epsilon = (t'' - t')/(N + 1)$, leads to the equation

$$\rho(x'', t'') = N' \int \cdots \int e^{-(1/2\nu\epsilon) \sum_{l=0}^N (x_{l+1} - x_l)^2} \prod_{l=1}^N dx_l \rho(x', t') dx' ,$$

where $x_{N+1} \equiv x''$ and $x_0 \equiv x'$. This equation features the imaginary time propagator for a free particle of unit mass as given formally as

$$W(x'', t''; x', t') = \mathcal{N} \int \mathcal{D}[x] e^{-(1/2\nu) \int \dot{x}^2 dt} ,$$

where \mathcal{N} denotes a formal normalization factor.

The similarity of this expression with the Feynman path integral [for $V(x) = 0$] is clear, but there is a profound difference between these equations. In the former (Feynman) case the underlying measure is only *finitely additive*, while in the latter (Wiener) case the continuum limit actually defines a genuine measure, i.e., a *countably additive measure* on paths, which is a version of the famous *Wiener measure*. In particular,

$$W(x'', t''; x', t') = \int d\mu_W'(x) ,$$

where μ_W' denotes a measure on continuous paths $x(t)$, $t' \leq t \leq t''$, for which $x(t'') \equiv x''$ and $x(t') \equiv x'$. Such a measure is said to be a *pinned* Wiener measure, since it specifies its path values at two time points, i.e., at $t = t'$ and at $t = t'' > t'$.

We note that Brownian motion paths have the property that with probability one they are concentrated on continuous paths. However, it is also true that the time derivative of a Brownian path is almost nowhere defined, which means that, with probability one, $\dot{x}(t) = \pm\infty$ for all t .

When the potential $V(x) \neq 0$ the propagator associated with (6.11) is formally given by

$$W(x'', t''; x', t') = \mathcal{N} \int \mathcal{D}[x] e^{-(1/2\nu) \int \dot{x}^2 dt - \int V(x) dt},$$

an expression which is well defined if $V(x) \geq c$, $-\infty < c < \infty$. A mathematically improved expression makes use of the Wiener measure and reads

$$W(x'', t''; x', t') = \int e^{-\int V(x(t)) dt} d\mu_W^\nu(x).$$

This is an elegant relation in that it represents a solution to the differential equation (6.11) in the form of an integral over Brownian motion paths suitably weighted by the potential V . Incidentally, since the propagator is evidently a strictly positive function, it follows that the solution of the differential equation (6.11) is nonnegative for all time t provided it is nonnegative for any particular time value.

6.2.5 Itô Formula

Itô [Ito (1960)] proposed another version of a *continuous-time regularization* that resolved some of the troublesome issues. In essence, the proposal of Itô takes the form given by

$$\lim_{\nu \rightarrow \infty} \mathcal{N}_\nu \int \mathcal{D}[x] \exp\{(i/\hbar) \int [\frac{1}{2} m \dot{x}^2 - V(x)] dt\} \exp\{-(1/2\nu) \int [\ddot{x}^2 + \dot{x}^2] dt\}.$$

Note well the alternative form of the auxiliary factor introduced as a regulator. The additional term \ddot{x}^2 , the square of the second derivative of x , acts to smooth out the paths sufficiently well so that in the case of (21) both $x(t)$ and $\dot{x}(t)$ are continuous functions, leaving $\ddot{x}(t)$ as the term which does not exist. However, since only x and \dot{x} appear in the rest of the integrand, the indicated path integral can be well defined; this is already a positive contribution all by itself (see e.g., [Klauder (1997); Klauder (2000)]).

6.3 Standard Path-Integral Quantization

6.3.1 Canonical versus Path-Integral Quantization

Recall that in the usual, *canonical formulation* of quantum mechanics, the system's phase-space coordinates, q , and momenta, p , are replaced by the

corresponding Hermitian operators in the Hilbert space, with real measurable eigenvalues, which obey *Heisenberg commutation relations*.

The *path-integral quantization* is instead based directly on the notion of a propagator $K(q_f, t_f; q_i, t_i)$ which is defined such that (see [Ryder (1996); Cheng and Li (1984); Gunion (2003)])

$$\psi(q_f, t_f) = \int K(q_f, t_f; q_i, t_i) \psi(q_i, t_i) dq_i, \quad (6.12)$$

i.e., the wave function $\psi(q_f, t_f)$ at final time t_f is given by a Huygens principle in terms of the wave function $\psi(q_i, t_i)$ at an initial time t_i , where we have to integrate over all the points q_i since all can, in principle, send out little wavelets that would influence the value of the wave function at q_f at the later time t_f . This equation is very general and is an expression of causality. We use the normal units with $\hbar = 1$.

According to the usual interpretation of quantum mechanics, $\psi(q_f, t_f)$ is the *probability amplitude* that the particle is at the point q_f and the time t_f , which means that $K(q_f, t_f; q_i, t_i)$ is the probability amplitude for a transition from q_i and t_i to q_f and t_f . The probability that the particle is observed at q_f at time t_f if it began at q_i at time t_i is

$$P(q_f, t_f; q_i, t_i) = |K(q_f, t_f; q_i, t_i)|^2.$$

Let us now divide the time interval between t_i and t_f into two, with t as the intermediate time, and q the intermediate point in space. Repeated application of (6.12) gives

$$\psi(q_f, t_f) = \int \int K(q_f, t_f; q, t) dq K(q, t; q_i, t_i) \psi(q_i, t_i) dq_i,$$

from which it follows that

$$K(q_f, t_f; q_i, t_i) = \int dq K(q_f, t_f; q, t) K(q, t; q_i, t_i).$$

This equation says that the transition from (q_i, t_i) to (q_f, t_f) may be regarded as the result of the transition from (q_i, t_i) to all available intermediate points q followed by a transition from (q, t) to (q_f, t_f) . This notion of *all possible paths* is crucial in the *path-integral formulation* of quantum mechanics.

Now, recall that the *state vector* $|\psi, t\rangle_S$ in the *Schrödinger picture* is related to that in the *Heisenberg picture* $|\psi\rangle_H$ by

$$|\psi, t\rangle_S = e^{-iHt} |\psi\rangle_H,$$

or, equivalently,

$$|\psi\rangle_H = e^{iHt} |\psi, t\rangle_S.$$

We also define the vector

$$|q, t\rangle_H = e^{iHt} |q\rangle_S,$$

which is the Heisenberg version of the Schrödinger state $|q\rangle$. Then, we can equally well write

$$\psi(q, t) = \langle q, t | \psi \rangle_H. \quad (6.13)$$

By completeness of states we can now write

$$\langle q_f, t_f | \psi \rangle_H = \int \langle q_f, t_f | q_i, t_i \rangle_H \langle q_i, t_i | \psi \rangle_H dq_i,$$

which with the definition of (6.13) becomes

$$\psi(q_f, t_f) = \int \langle q_f, t_f | q_i, t_i \rangle_H \psi(q_i, t_i) dq_i.$$

Comparing with (6.12), we get

$$K(q_f, t_f; q_i, t_i) = \langle q_f, t_f | q_i, t_i \rangle_H.$$

Now, let us calculate the *quantum-mechanics propagator*

$$\langle q', t' | q, t \rangle_H = \langle q' | e^{-iH(t-t')} | q \rangle$$

using the *path-integral formalism* that will incorporate the direct quantization of the coordinates, without Hilbert space and Hermitian operators.

The first step is to divide up the time interval into $n + 1$ tiny pieces: $t_l = l\varepsilon + t$ with $t' = (n + 1)\varepsilon + t$. Then, by completeness, we can write (dropping the Heisenberg picture index H from now on)

$$\begin{aligned} \langle q', t' | q, t \rangle &= \int dq_1(t_1) \dots \int dq_n(t_n) \langle q', t' | q_n, t_n \rangle \times \\ &\times \langle q_n, t_n | q_{n-1}, t_{n-1} \rangle \dots \langle q_1, t_1 | q, t \rangle. \end{aligned} \quad (6.14)$$

The integral $\int dq_1(t_1) \dots dq_n(t_n)$ is an *integral over all possible paths*, which are not trajectories in the normal sense, since there is no requirement of continuity, but rather *Markov chains*.

Now, for small ε we can write

$$\langle q', \varepsilon | q, 0 \rangle = \langle q' | e^{-i\varepsilon H(P, Q)} | q \rangle = \delta(q' - q) - i\varepsilon \langle q' | H(P, Q) | q \rangle,$$

where $H(P, Q)$ is the Hamiltonian (e.g., $H(P, Q) = \frac{1}{2}P^2 + V(Q)$, where P, Q are the momentum and coordinate operators). Then we have (see [Ryder (1996); Cheng and Li (1984); Gunion (2003)])

$$\langle q' | H(P, Q) | q \rangle = \int \frac{dp}{2\pi} e^{ip(q'-q)} H\left(p, \frac{1}{2}(q' + q)\right).$$

Putting this into our earlier form we get

$$\langle q', \varepsilon | q, 0 \rangle \simeq \int \frac{dp}{2\pi} \exp \left[i \left\{ p(q' - q) - \varepsilon H\left(p, \frac{1}{2}(q' + q)\right) \right\} \right],$$

where the 0th order in $\varepsilon \rightarrow \delta(q' - q)$ and the 1st order in $\varepsilon \rightarrow -i\varepsilon \langle q' | H(P, Q) | q \rangle$. If we now substitute many such forms into (6.14) we finally get

$$\begin{aligned} \langle q', t' | q, t \rangle &= \lim_{n \rightarrow \infty} \int \prod_{i=1}^n dq_i \prod_{k=1}^{n+1} \frac{dp_k}{2\pi} \times \\ &\times \exp \left\{ i \sum_{j=1}^{n+1} [p_j(q_j - q_{j-1})] - H\left(p_j, \frac{1}{2}(q_j + q_{j+1})\right) (t_j - t_{j-1}) \right\}, \end{aligned} \quad (6.15)$$

with $q_0 = q$ and $q_{n+1} = q'$. Roughly, the above formula says to *integrate over all possible momenta and coordinate values associated with a small interval*, weighted by something that is going to turn into the *exponential of the action* e^{iS} in the limit where $\varepsilon \rightarrow 0$. It should be stressed that the different q_i and p_k integrals are independent, which implies that p_k for one interval can be completely different from the $p_{k'}$ for some other interval (including the neighboring intervals). In principle, the integral (6.15) should be defined by *analytic continuation into the complex plane* of, for example, the p_k integrals.

Now, if we go to the differential limit where we call $t_j - t_{j-1} \equiv d\tau$ and write $\frac{(q_j - q_{j-1})}{(t_j - t_{j-1})} \equiv \dot{q}$, then the above formula takes the form

$$\langle q', t' | q, t \rangle = \int \mathcal{D}[p] \mathcal{D}[q] \exp \left\{ i \int_t^{t'} [p\dot{q} - H(p, q)] d\tau \right\},$$

where we have used the shorthand notation

$$\int \mathcal{D}[p] \mathcal{D}[q] \equiv \int \prod_{\tau} \frac{dq(\tau) dp(\tau)}{2\pi}.$$

Note that the above integration is an integration over the p and q values at every time τ . This is what we call a *functional integral*. We can think of a given set of choices for all the $p(\tau)$ and $q(\tau)$ as defining a *path in the 6D phase-space*. The most important point of the above result is that we have got an expression for a *quantum-mechanical transition amplitude* in terms of an integral involving only pure complex numbers, without operators.

We can actually perform the above integral for Hamiltonians of the type $H = H(P, Q)$. We use square completion in the exponential for this, defining the integral in the complex p plane and continuing to the physical situation. In particular, we have

$$\int_{-\infty}^{\infty} \frac{dp}{2\pi} \exp \left\{ i\varepsilon \left(p\dot{q} - \frac{1}{2}p^2 \right) \right\} = \frac{1}{\sqrt{2\pi i\varepsilon}} \exp \left[\frac{1}{2}i\varepsilon \dot{q}^2 \right],$$

(see [Ryder (1996); Cheng and Li (1984); Gunion (2003)]) which, substituting into (6.15) gives

$$\langle q', t' | q, t \rangle = \lim_{n \rightarrow \infty} \int \prod_i \frac{dq_i}{\sqrt{2\pi i\varepsilon}} \exp \left\{ i\varepsilon \sum_{j=1}^{n+1} \left[\frac{1}{2} \left(\frac{q_j - q_{j-1}}{\varepsilon} \right)^2 - V \left(\frac{q_j + q_{j+1}}{2} \right) \right] \right\}.$$

This can be formally written as

$$\langle q', t' | q, t \rangle = \int \mathcal{D}[q] e^{iS[q]},$$

where

$$\int \mathcal{D}[q] \equiv \int \prod_i \frac{dq_i}{\sqrt{2\pi i\varepsilon}},$$

while

$$S[q] = \int_t^{t'} L(q, \dot{q}) d\tau$$

is the *standard action* with the *Lagrangian*

$$L = \frac{1}{2} \dot{q}^2 - V(q).$$

Generalization to many degrees of freedom is straightforward:

$$\langle q_1' \dots q_N', t' | q_1 \dots q_N, t \rangle = \int \mathcal{D}[p] \mathcal{D}[q] \exp \left\{ i \int_t^{t'} \left[\sum_{n=1}^N p_n \dot{q}_n - H(p_n, q_n) \right] d\tau \right\},$$

with $\int \mathcal{D}[p] \mathcal{D}[q] = \int \prod_{n=1}^N \frac{dq_n dp_n}{2\pi}.$

Here, $q_n(t) = q_n$ and $q_n(t') = q_n'$ for all $n = 1, \dots, N$, and we are allowing for the full Hamiltonian of the system to depend upon all the N momenta and coordinates collectively.

6.3.2 Application: *Particles, Sources, Fields and Gauges*

6.3.2.1 *Particles*

(i) Consider first

$$\begin{aligned} & \langle q', t' | Q(t_0) | q, t \rangle \\ &= \int \prod dq_i(t_i) \langle q', t' | q_n, t_n \rangle \dots \langle q_{i0}, t_{i0} | Q(t_0) | q_{i-1}, t_{i-1} \rangle \dots \langle q_1, t_1 | q, t \rangle, \end{aligned}$$

where we choose one of the time interval ends to coincide with t_0 , i.e., $t_{i0} = t_0$. If we operate $Q(t_0)$ to the left, then it is replaced by its eigenvalue $q_{i0} = q(t_0)$. Aside from this one addition, everything else is evaluated just as before and we will obviously get

$$\langle q', t' | Q(t_0) | q, t \rangle = \int \mathcal{D}[p] \mathcal{D}[q] q(t_0) \exp \left\{ i \int_t^{t'} [p\dot{q} - H(p, q)] d\tau \right\}.$$

(ii) Next, suppose we want a *path-integral expression* for $\langle q', t' | Q(t_1) Q(t_2) | q, t \rangle$ in the case where $t_1 > t_2$. For this, we have to insert as intermediate states $|q_{i1}, t_{i1}\rangle \langle q_{i1}, t_{i1}|$ with $t_{i1} = t_1$ and $|q_{i2}, t_{i2}\rangle \langle q_{i2}, t_{i2}|$ with $t_{i2} = t_2$ and since we have ordered the times at which we do the insertions we must have the first insertion to the left of the 2nd insertion when $t_1 > t_2$. Once these insertions are done, we evaluate $\langle q_{i1}, t_{i1} | Q(t_1) = \langle q_{i1}, t_{i1} | q(t_1)$ and $\langle q_{i2}, t_{i2} | Q(t_2) = \langle q_{i2}, t_{i2} | q(t_2)$ and then proceed as before and get

$$\langle q', t' | Q(t_1) Q(t_2) | q, t \rangle = \int \mathcal{D}[p] \mathcal{D}[q] q(t_1) q(t_2) \exp \left\{ i \int_t^{t'} [p\dot{q} - H(p, q)] d\tau \right\}.$$

Now, let us ask what the above integral is equal to if $t_2 > t_1$? It is obvious that what we get for the above integral is $\langle q', t' | Q(t_2) Q(t_1) | q, t \rangle$. Clearly, this generalizes to an arbitrary number of Q operators.

(iii) When we enter into quantum field theory, the Q 's will be replaced by fields, since it is the fields that play the role of coordinates in the 2nd quantization conditions.

6.3.2.2 Sources

The *source* is represented by modifying the Lagrangian:

$$L \rightarrow L + J(t)q(t).$$

Let us define $|0, t\rangle^J$ as the ground state (vacuum) vector (in the moving frame, i.e., with the e^{iHt} included) in the presence of the source. The required *transition amplitude* is

$$Z[J] \propto \langle 0, +\infty | 0, -\infty \rangle^J,$$

where the source $J = J(t)$ plays a role analogous to that of an electromagnetic current, which acts as a source of the electromagnetic field. In other words, we can think of the scalar product $J_\mu A^\mu$, where J_μ is the current from a scalar (or Dirac) field acting as a source of the potential A^μ . In the same way, we can always define a current J that acts as the source for some arbitrary field ϕ . $Z[J]$ (otherwise denoted by $W[J]$) is a functional of the current J , defined as (see [Ryder (1996); Cheng and Li (1984); Gunion (2003)])

$$Z[J] \propto \int \mathcal{D}[p] \mathcal{D}[q] \exp \left\{ i \int_t^{t'} [p(\tau) \dot{q}(\tau) - H(p, q) + J(\tau)q(\tau)] d\tau \right\},$$

with the *normalization condition* $Z[J = 0] = 1$. Here, the argument of the exponential depends upon the functions $q(\tau)$ and $p(\tau)$ and we then integrate over all possible forms of these two functions. So the exponential is a functional that maps a choice for these two functions into a number. For example, for a quadratically completable $H(p, q)$, the p integral can be performed as a q integral

$$Z[J] \propto \int \mathcal{D}[q] \exp \left\{ i \int_{-\infty}^{+\infty} \left(L + Jq + \frac{1}{2} i \varepsilon q^2 \right) d\tau \right\},$$

where the addition to H was chosen in the form of a *convergence factor* $-\frac{1}{2} i \varepsilon q^2$.

6.3.2.3 Fields

Let us now treat the *abstract scalar field* $\phi(x)$ as a coordinate in the sense that we imagine dividing space up into many little cubes and the average value of the field $\phi(x)$ in that cube is treated as a coordinate for that little cube. Then, we go through the multi-coordinate analogue of the procedure we just considered above and take the continuum limit. The final result is

$$Z[J] \propto \int \mathcal{D}[\phi] \exp \left\{ i \int d^4x \left(\mathcal{L}(\phi(x)) + J(x)\phi(x) + \frac{1}{2}i\varepsilon\phi^2 \right) \right\},$$

where for \mathcal{L} we would employ the *Klein–Gordon Lagrangian* form. In the above, the dx_0 integral is the same as $d\tau$, while the $d^3\vec{x}$ integral is summing over the sub-Lagrangians of all the different little cubes of space and then taking the continuum limit. \mathcal{L} is the *Lagrangian density* describing the Lagrangian for each little cube after taking the many-cube limit (see [Ryder (1996); Cheng and Li (1984); Gunion (2003)]) for the full derivation).

We can now introduce *interactions*, \mathcal{L}_I . Assuming the simple form of the Hamiltonian, we have

$$Z[J] \propto \int \mathcal{D}[\phi] \exp \left\{ i \int d^4x \left(\mathcal{L}(\phi(x)) + \mathcal{L}_I(\phi(x)) + J(x)\phi(x) \right) \right\},$$

again using the normalization factor required for $Z[J=0] = 1$.

For example of Klein Gordon theory, we would use

$$\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_I, \quad \mathcal{L}_0 = \frac{1}{2}[\partial_\mu\phi\partial^\mu\phi - \mu^2\phi^2], \quad \mathcal{L}_I = \mathcal{L}_I(\phi),$$

where $\partial_\mu \equiv \partial_{x^\mu}$ and we can freely manipulate indices, as we are working in Euclidean space \mathbb{R}^3 . In order to define the above $Z[J]$, we have to include a convergence factor $i\varepsilon\phi^2$,

$$\mathcal{L}_0 \rightarrow \frac{1}{2}[\partial_\mu\phi\partial^\mu\phi - \mu^2\phi^2 + i\varepsilon\phi^2], \quad \text{so that}$$

$$Z[J] \propto \int \mathcal{D}[\phi] \exp \left\{ i \int d^4x \left(\frac{1}{2}[\partial_\mu\phi\partial^\mu\phi - \mu^2\phi^2 + i\varepsilon\phi^2] + \mathcal{L}_I(\phi(x)) + J(x)\phi(x) \right) \right\}$$

is the appropriate *generating function* in the free field theory case.

6.3.2.4 Gauges

In the path integral approach to quantization of the *gauge theory*, we implement *gauge fixing* by restricting in some manner or other the path integral

over gauge fields $\int \mathcal{D}[A_\mu]$. In other words we will write instead

$$Z[J] \propto \int \mathcal{D}[A_\mu] \delta(\text{some gauge fixing condition}) \exp\{i \int d^4x \mathcal{L}(A_\mu)\}.$$

A common approach would be to start with the *gauge condition*

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{2}(\partial^\mu A_\mu)^2$$

where the electrodynamic field tensor is given by $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, and calculate

$$Z[J] \propto \int \mathcal{D}[A_\mu] \exp\left\{i \int d^4x [\mathcal{L}(A_\mu(x)) + J_\mu(x)A^\mu(x)]\right\}$$

as the *generating function* for the *vacuum expectation* values of *time ordered products* of the A_μ fields. Note that J_μ should be conserved ($\partial^\mu J_\mu = 0$) in order for the full expression $\mathcal{L}(A_\mu) + J_\mu A^\mu$ to be *gauge-invariant* under the integral sign when $A_\mu \rightarrow A_\mu + \partial^\mu \Lambda$. For a proper approach, see [Ryder (1996); Cheng and Li (1984); Gunion (2003)].

6.3.3 Riemannian–Symplectic Geometries

In this section, following [Shabanov and Klauder (1998)], we describe path integral quantization on Riemannian–symplectic manifolds. Let \hat{q}^j be a set of Cartesian coordinate canonical operators satisfying the Heisenberg commutation relations $[\hat{q}^j, \hat{q}^k] = i\omega^{jk}$. Here $\omega^{jk} = -\omega^{kj}$ is the canonical symplectic structure. We introduce the canonical coherent states as $|q\rangle \equiv e^{iq^j \omega_{jk} \hat{q}^k} |0\rangle$, where $\omega_{jn} \omega^{nk} = \delta_j^k$, and $|0\rangle$ is the ground state of a harmonic oscillator with unit angular frequency. Any state $|\psi\rangle$ is given as a function on phase-space in this representation by $\langle q|\psi\rangle = \psi(q)$. A general operator \hat{A} can be represented in the form $\hat{A} = \int dq a(q)|q\rangle\langle q|$, where $a(q)$ is the lower symbol of the operator and dq is a properly normalized form of the Liouville measure. The function $A(q, q') = \langle q|\hat{A}|q'\rangle$ is the *kernel of the operator*.

The main object of the path integral formalism is the integral kernel of the evolution operator

$$K_t(q, q') = \langle q|e^{-it\hat{H}}|q'\rangle = \int_{q(0)=q'}^{q(t)=q} \mathcal{D}[q] e^{i \int_0^t d\tau (\frac{1}{2} \dot{q}^j \omega_{jk} \dot{q}^k - h)}. \quad (6.16)$$

Here \hat{H} is the Hamiltonian, and $h(q)$ its symbol. The measure formally implies a sum over all phase-space paths pinned at the initial and final points, and a Wiener measure regularization implies the following replacement

$$\mathcal{D}[q] \rightarrow \mathcal{D}[\mu_\nu(q)] = \mathcal{D}[q] e^{-\frac{1}{2\nu} \int_0^t d\tau \dot{q}^2} = N_\nu(t) d\mu_W^\nu(q) . \quad (6.17)$$

The factor $N_\nu(t)$ equals $2\pi e^{\nu t/2}$ for every degree of freedom, $d\mu_W^\nu(q)$ stands for the Wiener measure, and ν denotes the diffusion constant. We denote by $K_t^\nu(q, q')$ the integral kernel of the evolution operator for a finite ν . The Wiener measure determines a stochastic process on the *flat* phase-space. The integral of the symplectic 1-form $\int q\omega dq$ is a stochastic integral that is interpreted in the Stratonovich sense. Under general coordinate transformations $q = q(\bar{q})$, the Wiener measure describes *the same* stochastic process on *flat* space in the curvilinear coordinates $dq^2 = d\sigma(\bar{q})^2$, so that the value of the integral is not changed apart from a possible phase term. After the calculation of the integral, the evolution operator kernel is get by taking the limit $\nu \rightarrow \infty$. The existence of this limit, and also the covariance under general phase-space coordinate transformations, can be proved through the *operator* formalism for the regularized kernel $K_t^\nu(q, q')$.

Note that the integral (6.16) with the Wiener measure inserted can be regarded as an ordinary Lagrangian path integral with a complex action, where the configuration space is the original phase-space and the Hamiltonian $h(q)$ serves as a potential. Making use of this observation it is not hard to derive the corresponding Schrödinger-like equation

$$\partial_t K_t^\nu(q, q') = \left[\frac{\nu}{2} \left(\partial_{q^j} + \frac{i}{2} \omega_{jk} q^k \right)^2 - i h(q) \right] K_t^\nu(q, q') , \quad (6.18)$$

subject to the initial condition $K_{t=0}^\nu(q, q') = \delta(q - q')$, $0 < \nu < \infty$. One can show that $\hat{K}_t^\nu \rightarrow \hat{K}_t$ as $\nu \rightarrow \infty$ for all $t > 0$. The covariance under general coordinate transformations follows from the covariance of the ‘kinetic’ energy of the Schrödinger operator in (6.18): The Laplace operator is replaced by the Laplace–Beltrami operator in the new curvilinear coordinates $q = q(\bar{q})$, so the solution is not changed, but written in the new coordinates. This is similar to the covariance of the ordinary Schrödinger equation and the corresponding *Lagrangian* path integral relative to general coordinate transformations on the configuration space: The kinetic energy operator (the Laplace operator) in the ordinary Schrödinger equation gives a term *quadratic* in time derivatives in the path integral measure which is sufficient for the general coordinate covariance. We remark that the regularization

procedure based on the modified Schrödinger equation (6.18) applies to far more general Hamiltonians than those quadratic in canonical momenta and leading to the conventional *Lagrangian* path integral.

6.3.4 Euclidean Stochastic Path Integral

Recall that modern stochastic calculus permits development of three alternative descriptions of complex stochastic systems:

- (1) Langevin rate ODEs [Gardiner (1985)],
- (2) Fokker–Planck PDEs [Gardiner (1985)], and
- (3) Euclidean stochastic path–integrals [Wehner and Wolfer (1983a); Wehner and Wolfer (1983b); Graham (1978); Langouche *et. al.* (1980)].

L. Ingber [Ingber (1997)] has successfully applied a stochastic path–integral of the Euclidean form

$$\langle f|i \rangle = \int_{\Omega} \mathcal{D}[w] e^{-\mathcal{A}[w]},$$

to in–depth analysis of three completely different *non–equilibrium nonlinear multivariate* Gaussian–Markovian systems:

- (1) Statistical mechanical descriptions of neocortex (short–term memory and EEG),
- (2) Analysis of financial markets (interest–rate and trading models), and
- (3) Combat analysis (baselining simulations to exercise data).

The core of all his work are of context–specific *mesoscopic order parameters* M^G . Ingber starts with Langevin rate equations (with zero–mean Gaussian white noise $\eta^j(t)$), as *order parameter equations* (see [Haken (1983); Haken (1993)])

$$\begin{aligned} \dot{M}^G &= f^G + \hat{g}_j^G \eta^j, \quad (G = 1, \dots, \Lambda), \quad (j = 1, \dots, N), \\ < \eta^j(t) >_{\eta} = 0, < \eta^j(t), \eta^{j'}(t') >_{\eta} = \delta^{jj'} \delta(t - t'), \end{aligned} \quad (6.19)$$

where f^G and \hat{g}_j^G are generally nonlinear functions of mesoscopic order parameters M^G , j is a microscopic index indicating the source of fluctuations, and $N \geq \Lambda$.

Via a somewhat lengthy calculation, involving an intermediate derivation of a corresponding Fokker–Planck or Schrödinger–type equation for the conditional probability distribution $P[M(t)|M(t_0)]$, equation (6.19) is

developed into the more useful probability distribution for the order parameters M^G at long-time macroscopic time event $t = (u + 1)\theta + t_0$, in terms of a Stratonovich–Riemannian path-integral over mesoscopic Gaussian conditional probabilities. Here, macroscopic variables are defined as the long-time limit of the evolving mesoscopic system.

The corresponding Schrödinger-type equation is

$$\begin{aligned}\partial_t P &= \frac{1}{2}(g^{GG'}P)_{,GG'} - (g^G P)_{,G} + V, \\ g^{GG'} &= k_T \delta^{jk} \hat{g}_j^G \hat{g}_k^{G'}, \\ g^G &= f^G + \frac{1}{2} \delta^{jk} \hat{g}_j^{G'} \hat{g}_{k,G'}^G, \\ [.]_{,G} &= \partial[.]/\partial M^G.\end{aligned}\tag{6.20}$$

This is properly referred to as a Fokker–Planck equation when $V = 0$. Note that although the partial differential equation (6.20) contains equivalent information regarding the order parameters M^G as in the stochastic differential equation (6.19), all references to j have been properly averaged over, i.e., \hat{g}_j^G in (6.19) is an entity with parameters in both microscopic and mesoscopic spaces, but M is a purely mesoscopic variable, and this is more clearly reflected in (6.20).

Now, the path integral representation is given in terms of the Lagrangian $L = L(\dot{M}^G, M^G, t)$, as

$$P[M_t | M_{t_0}] dM(t) = \int \mathcal{D}[M] \exp(-A[M]) \delta[M(t_0) = M_0] \delta[M(t) = M_t],\tag{6.21}$$

where the action $A[M]$ is given by

$$A[M] = k_T^{-1} \min \int_{t_0}^t dt' L(\dot{M}^G, M^G, t),$$

and the *path measure* $\mathcal{D}[M]$ is equal to

$$\mathcal{D}[M] = \lim_{u \rightarrow \infty} \prod_{\rho=1}^{u+1} g^{1/2} \prod_G (2\pi\theta)^{-1/2} dM_\rho^G.$$

The Lagrangian L is given by

$$L(\dot{M}^G, M^G, t) = \frac{1}{2}(\dot{M}^G - h^G)g_{GG'}(\dot{M}^{G'} - h^{G'}) + \frac{1}{2}h_{,G}^G + R/6 - V,$$

where $h^G = g^G - \frac{1}{2}g^{-1/2}(g^{1/2}g^{GG'})_{,G'}$, comma ‘,’ denotes partial derivative, and $g_{GG'}$ is the inverse metric tensor $g_{GG'} = (g^{GG'})^{-1}$, with determinant $g = \det(g_{GG'})$. The covariant derivative $h^G_{;G}$ is defined classically through the Christoffel symbols Γ^F_{GF} ,

$$\begin{aligned} h^G_{;G} &= h^G_{,G} + \Gamma^F_{GF} h^G = g^{-1/2}(g^{1/2}h^G)_{,G}, \\ \Gamma^F_{GF} &= g^{LF}(g_{JL,K} + g_{KL,J} - g_{JK,L}), \end{aligned}$$

while the Riemannian curvature is defined as

$$\begin{aligned} R &= g^{JL}R_{JL} = g^{JL}g^{JK}R_{FJKL}, \\ R_{FJKL} &= \frac{1}{2}(g_{FK,JL} - g_{JK,FL} - g_{FL,JK} + g_{JL,FK}) + g_{MN}(\Gamma^M_{FK}\Gamma^N_{FL} - \Gamma^N_{JK}). \end{aligned}$$

Therefore, mesoscopic variables have been defined as M^G in the Langevin and Fokker–Planck representations, in terms of their development from the microscopic system labelled by j . The Riemannian curvature term R arises from nonlinear $g_{GG'}$, which is a genuine metric of this space [Graham (1978)]. Even if a stationary solution, i.e., $\dot{M}^G = 0$, is ultimately sought, a necessarily prior stochastic treatment of \dot{M}^G terms gives rise to these Riemannian ‘corrections’. Even for a constant metric, the term $h^G_{;G}$ contributes to L for a nonlinear mean h^G . V may include terms such as $\sum_{T'} J_{T'G} M^G$, where the Lagrange multipliers $J_{T'G}$ are constraints on M^G , which are advantageously modelled as extrinsic sources in this representation; they too may be time-dependent. Using the variational principle, J_{TG} may also be used to constrain M^G to regions where they are empirically bound. With respect to a steady state \bar{P} , when it exists, the *information gain* in state P is defined by

$$I[P] = \int \mathcal{D}[M'] P \ln(P/\bar{P}), \quad \text{where} \quad \mathcal{D}[M'] = \mathcal{D}[M]/dM_{u+1}.$$

A robust and accurate histogram-based (i.e., non-Monte Carlo) path-integral algorithm to calculate the long-time probability distribution has been developed to handle nonlinear Lagrangians (see [Wehner and Wolfer (1983a); Wehner and Wolfer (1983b); Graham (1978); Langouche *et al.* (1980)]).

In the economics literature, there appears to be sentiment to define (6.19) by the Ito, rather than the Stratonovich prescription. It should be noted that virtually all investigations of other physical systems, which are also continuous time models of discrete processes, conclude that the Stratonovich interpretation coincides with reality, when multiplicative noise

with zero correlation time, modelled in terms of white noise η^j , is properly considered as the limit of real noise with finite correlation time (see [Gardiner (1985)]). The path integral succinctly demonstrates the difference between the two: The *Ito prescription* corresponds to the so-called *prepoint discretization* of L , wherein

$$\theta \dot{M}(t) \rightarrow M_{\rho+1} - M_{\rho} \quad \text{and} \quad M(t) \rightarrow M_{\rho}.$$

The *Stratonovich prescription* corresponds to the *midpoint discretization* of L , wherein

$$\theta \dot{M}(t) \rightarrow M_{\rho+1} - M_{\rho} \quad \text{and} \quad M(t) \rightarrow \frac{1}{2}(M_{\rho+1} + M_{\rho}).$$

In terms of the functions appearing in the Fokker–Planck equation (6.20), the Ito prescription of the prepoint discretized Lagrangian, L_I , is relatively simple:

$$L_I(\dot{M}^G, M^G, t) = \frac{1}{2}(\dot{M}^G - g^G)g_{GG'}(\dot{M}^{G'} - g^{G'}) - V,$$

however, this is deceptive because of its nonstandard calculus [Ingber (1997)].

Now, as L possesses a variational principle, sets of contour graphs, at different long-time epochs of the path-integral of P over its variables at all intermediate times, give a visually intuitive and accurate decision-aid to view the dynamic evolution of the scenario. For example, this Lagrangian approach permits a quantitative assessment of the following concepts usually only loosely defined:

$$\begin{aligned} \text{'Momentum'} &= \Pi^G = L_{\dot{M}^G} \\ \text{'Mass'} &= g_{GG'} = L_{\dot{M}^G \dot{M}^{G'}} \\ \text{'Force'} &= F = L_{M^G} \\ \text{'F} = \text{ma'} : 0 &= \delta L = L_{M^G} - \partial_t L_{\dot{M}^G} \end{aligned}$$

These physical entities provide another form of intuitive, but quantitatively precise, presentation of these analyzes. For example, daily newspapers use this terminology to discuss the movement of security prices. Π^G serve as *canonical momenta indicators* (CMI) for these systems [Ingber (1997)].

6.3.5 Application: *Stochastic Optimal Control*

A path–integral based optimal control model for nonlinear stochastic systems has recently been developed in [Kappen (2006)]. The author addressed the role of noise and the issue of efficient computation in stochastic optimal control problems. He considered a class of nonlinear control problems that can be formulated as a *path integral* and where the noise plays the role of temperature. The path integral displays symmetry breaking and there exist a critical noise value that separates regimes where optimal control yields qualitatively different solutions. The path integral can be computed efficiently by Monte Carlo integration or by Laplace approximation, and can therefore be used to solve high dimensional stochastic control problems.

Recall that *optimal control of nonlinear systems in the presence of noise* is a very general problem that occurs in many areas of science and engineering. It underlies *autonomous system behavior*, such as the control of movement and planning of actions of animals and robots, but also optimization of financial investment policies and control of chemical plants. The problem is stated as: *given that the system is in this configuration at this time, what is the optimal course of action to reach a goal state at some future time*. The *cost* of each time course of actions consists typically of a path contribution, that specifies the amount of work or other cost of the trajectory, and an end cost, that specifies to what extend the trajectory reaches the goal state.

Also recall that in the absence of noise, the optimal control problem can be solved in two ways: using (i) the *Pontryagin Maximum Principle* (PMP, see previous subsection), which represents a pair of ordinary differential equations that are similar to the Hamiltonian equations; or (ii) the *Hamilton–Jacobi–Bellman* (HJB) equation, which is a partial differential equation (PDE) [Bellman and Kalaba (1964)].

In the presence of *Wiener noise*, the PMP formalism is replaced by a set of stochastic differential equations (SDEs), which become difficult to solve (compare with [Yong and Zhou (1999)]). The inclusion of noise in the HJB framework is mathematically quite straightforward, yielding the so–called *stochastic HJB equation* [Stengel (1993)]. However, its solution requires a *discretization of space and time* and the computation becomes intractable in both memory requirement and CPU time in high dimensions. As a result, deterministic control can be computed efficiently using the PMP approach, but stochastic control is intractable due to the curse of dimensionality.

For small noise, one expects that optimal stochastic control resembles

optimal deterministic control, but for larger noise, the optimal stochastic control can be entirely different from the deterministic control [Russell and Norvig (2003)]. However, there is currently no good understanding how *noise* affects optimal control.

In this subsection, we address both the issue of efficient computation and the role of noise in stochastic optimal control. We consider a class of nonlinear stochastic control problems, that can be formulated as a statistical mechanics problem. This class of control problems includes arbitrary dynamical systems, but with a limited control mechanism. It contains linear-quadratic [Stengel (1993)] control as a special case. We show that under certain conditions on the noise, the HJB equation can be written as a *linear* PDE

$$-\partial_t \psi = H\psi, \quad (6.22)$$

with H a (non-Hermitian) operator. Equation (6.22) must be solved subject to a boundary condition at the end time. As a result of the linearity of (6.22), the solution can be obtained in terms of a diffusion process evolving forward in time, and can be written as a path integral. The path-integral has a direct interpretation as a free energy, where noise plays the role of temperature.

This link between stochastic optimal control and a free energy has an immediate consequence that phenomena that allow for a free energy description, typically display phase transitions. [Kappen (2006)] has argued that for stochastic optimal control one can identify a critical noise value that separates regimes where the optimal control has been qualitatively different. He showed how the Laplace approximation can be combined with Monte Carlo sampling to efficiently calculate the optimal control.

6.3.5.1 Path-Integral Formalism

Let x^i be an nD stochastic variable that is subject to the SDE

$$dx^i = (b^i(x^i, t) + u^i)dt + d\xi^i \quad (6.23)$$

with $d\xi^i$ being an nD Wiener process with $\langle d\xi_i d\xi_j \rangle = \nu_{ij} dt$, and functions ν_{ij} independent of x^i, u^i and time t . The term $b^i(x^i, t)$ is an arbitrary nD function of x^i and t , and u^i represents an nD vector of control variables. Given the value of x^i at an initial time t , the *stochastic optimal control*

problem is to find the control path $u^i(\cdot)$ that minimizes

$$C(x^i, t, u^i(\cdot)) = \left\langle \phi(x^i(t_f)) + \int_t^{t_f} d\tau \left(\frac{1}{2} u_i(\tau) R u_i(\tau) + V(x^i(\tau), \tau) \right) \right\rangle_{x^i}, \quad (6.24)$$

with R a matrix, $V(x^i, t)$ a time-dependent potential, and $\phi(x^i)$ the end cost. The brackets $\langle \rangle_{x^i}$ denote expectation value with respect to the stochastic trajectories (6.23) that start at x^i .

One defines the *optimal cost-to-go function* from any time t and state x^i as

$$J(x^i, t) = \min_{u^i(\cdot)} C(x^i, t, u^i(\cdot)).$$

J satisfies the following stochastic HJB equation [Kappen (2006)]

$$\begin{aligned} -\partial_t J(x^i, t) &= \min_{u^i} \left(\frac{1}{2} u_i R u_i + V + (b_i + u_i) \partial_{x^i} J(x^i, t) + \frac{1}{2} \nu_{ij} \partial_{x^i x^j} J(x^i, t) \right) \\ &= -\frac{1}{2} R^{-1} \partial_{x^i} J(x^i, t) \partial_{x^i} J + V + b_i \partial_{x^i} J(x^i, t) + \frac{1}{2} \nu_{ij} \partial_{x^i x^j} J(x^i, t), \end{aligned} \quad (6.25)$$

where $b_i = (b^i)^T$, and $u_i = (u^i)^T$, and

$$u^i = -R^{-1} \partial_{x^i} J(x^i, t) \quad (6.26)$$

is the optimal control at the point (x^i, t) . The HJB equation is nonlinear in J and must be solved with end boundary condition $J(x^i, t_f) = \phi(x^i)$.

Let us define $\psi(x^i, t)$ through the Log Transform

$$J(x^i, t) = -\lambda \log \psi(x^i, t), \quad (6.27)$$

and assume that there exists a scalar λ such that

$$\lambda \delta_{ij} = (R\nu)_{ij}, \quad (6.28)$$

with δ_{ij} the Kronecker delta. In the one dimensional case, such a λ can always be found. In the higher dimensional case, this restricts the matrices $R \propto (\nu_{ij})^{-1}$. Equation (6.28) reduces the dependence of optimal control on the n D noise matrix to a scalar value λ that will play the role of temperature, while (6.25) reduces to the linear equation (6.22) with

$$H = -\frac{V}{\lambda} + b_i \partial_{x^i} + \frac{1}{2} \nu_{ij} \partial_{x^i x^j} J(x^i, t).$$

Let $\rho(y^i, \tau | x^i, t)$ with $\rho(y^i, t | x^i, t) = \delta(y^i - x^i)$ describe a *diffusion process* for $\tau > t$ defined by the *Fokker-Planck equation*

$$\partial_\tau \rho = H^\dagger \rho = -\frac{V}{\lambda} \rho - \partial_{x^i} (b_i \rho) + \frac{1}{2} \nu_{ij} \partial_{x^i x^j} J(x^i, t) \rho \quad (6.29)$$

with H^\dagger the Hermitian-conjugate of H . Then $A(\tau) = \int dy^i \rho(y^i, \tau | x^i, t) \psi(y^i, \tau)$ is independent of τ and in particular $A(t) = A(t_f)$. It immediately follows that

$$\psi(x^i, t) = \int dy^i \rho(y^i, t_f | x^i, t) \exp(-\phi(y^i)/\lambda) \quad (6.30)$$

We arrive at the important conclusion that $\psi(x^i, t)$ can be computed either by backward integration using (6.22) or by forward integration of a diffusion process given by (6.29).

We can write the integral in (6.30) as a path integral. Following [Kappen (2006)] we can divide the time interval $t \rightarrow t_f$ in n_1 intervals and write $\rho(y^i, t_f | x^i, t) = \prod_{i=1}^{n_1} \rho(x_i^i, t_i | x_{i-1}^i, t_{i-1})$ and let $n_1 \rightarrow \infty$. The result is

$$\psi(x^i, t) = \int [dx^i]_{x^i} \exp\left(-\frac{1}{\lambda} S(x^i(t \rightarrow t_f))\right) \quad (6.31)$$

with $\int [dx^i]_{x^i}$ an integral over all paths $x^i(t \rightarrow t_f)$ that start at x^i and with

$$S(x^i(t \rightarrow t_f)) = \phi(x^i(t_f)) + \int_t^{t_f} d\tau \left(\frac{1}{2} (\dot{x}^i - b_i(x^i, \tau)) R(\dot{x}^i - b^i(x^i, \tau)) + V(x^i, \tau) \right) \quad (6.32)$$

the Action associated with a path. From (6.27) and (6.31), the cost-to-go $J(x, t)$ becomes a log partition sum (i.e., a free energy) with temperature λ .

6.3.5.2 Monte Carlo Sampling

The path integral (6.31) can be estimated by stochastic integration from t to t_f of the diffusion process (6.29) in which particles get annihilated at a rate $V(x^i, t)/\lambda$ [Kappen (2006)]:

$$\begin{aligned} x^i &= x^i + b^i(x^i, t) dt + d\xi^i, & \text{with probability } 1 - V dt/\lambda \\ x^i &= \dagger, & \text{with probability } V dt/\lambda \end{aligned} \quad (6.33)$$

where \dagger denotes that the particle is taken out of the simulation. Denote the trajectories by $x_\alpha^i(t \rightarrow t_f)$, ($\alpha = 1, \dots, N$). Then, $\psi(x^i, t)$ and u^i are

estimated as

$$\hat{\psi}(x^i, t) = \sum_{\alpha \in \text{alive}} w_{\alpha}, \quad u^i dt = \frac{1}{\hat{\psi}(x^i, t)} \sum_{\alpha \in \text{alive}}^N w_{\alpha} d\xi_{\alpha}^i(t), \quad (6.34)$$

$$\text{with} \quad w_{\alpha} = \frac{1}{N} \exp(-\phi(x_{\alpha}^i(t_f))/\lambda),$$

where ‘alive’ denotes the subset of trajectories that do not get killed along the way by the \dagger operation. The normalization $1/N$ ensures that the annihilation process is properly taken into account. Equation (6.34) states that optimal control at time t is obtained by averaging the initial directions of the noise component of the trajectories $d\xi_{\alpha}^i(t)$, weighted by their success at t_f .

The above sampling procedure can be quite inefficient, when many trajectories get annihilated. One of the simplest procedures to improve it is by importance sampling. We replace the diffusion process that yields $\rho(y^i, t_f | x^i, t)$ by another diffusion process, that will yield $\rho'(y^i, t_f | x^i, t) = \exp(-S'/\lambda)$. Then (6.31) becomes,

$$\psi(x^i, t) = \int [dx^i]_{x^i} \exp(-S'/\lambda) \exp(-(S - S')/\lambda).$$

The idea is to chose ρ' such as to make the sampling of the path integral as efficient as possible. Following [Kappen (2006)], here we use the Laplace approximation, which is given by the k deterministic trajectories $x_{\beta}(t \rightarrow t_f)$ that minimize the Action

$$J(x^i, t) \approx -\lambda \log \sum_{\beta=1}^k \exp(-S(x_{\beta}^i(t \rightarrow t_f))/\lambda).$$

The Laplace approximation ignores all fluctuations around the modes and becomes exact in the limit $\lambda \rightarrow 0$. The Laplace approximation can be computed efficiently, requiring $\mathcal{O}(n^2 m^2)$ operations, where m is the number of time discretization.

For each Laplace trajectory, we can define a diffusion processes ρ'_{β} according to (6.33) with $b^i(x^i, t) = x_{\beta}^i(t)$. The estimators for ψ and u^i are given again by (6.34), but with weights

$$w_{\alpha} = \frac{1}{N} \exp(-(S(x_{\alpha}^i(t \rightarrow t_f)) - S'_{\beta}(x_{\alpha}^i(t \rightarrow t_f)))/\lambda).$$

S is the original Action (6.32) and S'_{β} is the new Action for the Laplace guided diffusion. When there are multiple Laplace trajectories one should

include all of these in the sample.

6.3.6 Application: *Nonlinear Dynamics of Option Pricing*

Classical theory of option pricing is based on the results found in 1973 by Black and Scholes [Black and Scholes (1973)] and, independently, Merton [Merton (1973)]. Their pioneering work starts from the basic assumption that the asset prices follow the dynamics of a particular stochastic process (geometrical Brownian motion), so that they have a lognormal distribution [Hull (2000); Paul and Baschnagel (1999)]. In the case of an efficient market with no arbitrage possibilities, no dividends and constant volatilities, they found that the price of each financial derivative is ruled by an ordinary partial differential equation, known as the (Nobel–Prize winning) *Black–Scholes–Merton* (BSM) formula. In the most simple case of a so-called European option, the BSM equation can be explicitly solved to get an analytical formula for the price of the option [Hull (2000); Paul and Baschnagel (1999)]. When we consider other financial derivatives, which are commonly traded in real markets and allow anticipated exercise and/or depend on the history of the underlying asset, the BSM formula fails to give an analytical result. Appropriate numerical procedures have been developed in the literature to price exotic financial derivatives with path-dependent features, as discussed in detail in [Hull (2000); Wilmott *et al.* (1993); Potters *et al.* (2001)]. The aim of this work is to give a contribution to the problem of efficient option pricing in financial analysis, showing how it is possible to use path integral methods to develop a fast and precise algorithm for the evaluation of option prices.

Following recent studies on the application of the path integral approach to the financial market as appeared in the econophysics literature (see [Matacz (2002)] for a comprehensive list of references), in [Montagna *et al.* (2002)] the authors proposed an original, efficient path integral algorithm to price financial derivatives, including those with path-dependent and early exercise features, and to compare the results with those get with the standard procedures known in the literature.

6.3.6.1 *Theory and Simulations of Option Pricing*

Classical Theory and Path-Dependent Options

The basic ingredient for the development of a theory of option pricing is

a suitable model for the time evolution of the asset prices. The assumption of the BSM model is that the price S of an asset is driven by a *Brownian motion* and verifies the stochastic differential equation (SDE) [Hull (2000); Paul and Baschnagel (1999)]

$$dS = \mu S dt + \sigma S dw, \quad (6.35)$$

which, by means of the *Itô lemma*, can be cast in the form of an arithmetic Brownian motion for the logarithm of S

$$d(\ln S) = A dt + \sigma dw, \quad (6.36)$$

where σ is the *volatility*, $A = (\mu - \sigma^2/2)$, μ is the drift parameter and w is the realization of a Wiener process. Due to the properties of a Wiener process, (6.36) may be written as

$$d(\ln S) = A dt + \sigma \epsilon \sqrt{dt}, \quad (6.37)$$

where ϵ follows from a standardized normal distribution with mean 0 and variance 1. Thus, in terms of the logarithms of the asset prices $z' = \ln S'$, $z = \ln S$, the conditional transition probability $p(z'|z)$ to have at the time t' a price S' under the hypothesis that the price was S at the time $t < t'$ is given by [Paul and Baschnagel (1999); Bennati *et. al.* (1999)]

$$p(z'|z) = \frac{1}{\sqrt{2\pi(t'-t)\sigma^2}} \exp \left\{ -\frac{[z' - (z + A(t'-t))]^2}{2\sigma^2(t'-t)} \right\}, \quad (6.38)$$

which is a gaussian distribution with mean $z + A(t' - t)$ and variance $\sigma^2(t' - t)$. If we require the options to be exercised only at specific times t_i , $i = 1, \dots, n$, the asset price, between two consequent times t_{i-1} and t_i , will follow (6.37) and the related transition probability will be

$$p(z_i|z_{i-1}) = \frac{1}{\sqrt{2\pi\Delta t\sigma^2}} \exp \left\{ -\frac{[z_i - (z_{i-1} + A\Delta t)]^2}{2\sigma^2\Delta t} \right\}, \quad (6.39)$$

with $\Delta t = t_i - t_{i-1}$.

A time-evolution model for the asset price is strictly necessary in a theory of option pricing because the fair price at time $t = 0$ of an option \mathcal{O} , without possibility of anticipated exercise before the expiration date or maturity T (a so-called *European option*), is given by the scaled expectation

value [Hull (2000)]

$$\mathcal{O}(0) = e^{-rT} E[\mathcal{O}(T)], \quad (6.40)$$

where r is the risk-free interest and $E[\cdot]$ indicates the mean value, which can be computed only if a model for the asset underlying the option is understood. For example, the value \mathcal{O} of an European call option at the maturity T will be $\max\{S_T - X, 0\}$, where X is the strike price, while for an European put option the value \mathcal{O} at the maturity will be $\max\{X - S_T, 0\}$. It is worth emphasizing, for what follows, that the case of an European option is particularly simple, since in such a situation the price of the option can be evaluated by means of analytical formulae, which are get by solving the BSM partial differential equation with the appropriate boundary conditions [Hull (2000); Paul and Baschnagel (1999)]. On the other hand, many further kinds of options are present in the financial markets, such as American options (options which can be exercised at any time up to the expiration date) and exotic options [Hull (2000)], i.e., derivatives with complicated payoffs or whose value depend on the whole time evolution of the underlying asset and not just on its value at the end. For such options with path-dependent and early exercise features no exact solutions are available and pricing them correctly is a great challenge.

In the case of options with possibility of anticipated exercise before the expiration date, the above discussion needs to be generalized, by introducing a slicing of the time interval T . Let us consider, for definiteness, the case of an option which can be exercised within the maturity but only at the times $t_1 = \Delta t, t_2 = 2\Delta t, \dots, t_n = n\Delta t = T$. At each time slice t_{i-1} the value \mathcal{O}_{i-1} of the option will be the maximum between its expectation value at the time t_i scaled with $e^{-r\Delta t}$ and its value in the case of anticipated exercise \mathcal{O}_{i-1}^Y . If S_{i-1} denotes the price of the underlying asset at the time t_{i-1} , we can thus write for each $i = 1, \dots, n$

$$\mathcal{O}_{i-1}(S_{i-1}) = \max \{ \mathcal{O}_{i-1}^Y(S_{i-1}), e^{-r\Delta t} E[\mathcal{O}_i | S_{i-1}] \}, \quad (6.41)$$

where $E[\mathcal{O}_i | S_{i-1}]$ is the conditional expectation value of \mathcal{O}_i , i.e., its expectation value under the hypothesis of having the price S_{i-1} at the time t_{i-1} . In this way, to get the actual price \mathcal{O}_0 , it is necessary to proceed backward in time and calculate $\mathcal{O}_{n-1}, \dots, \mathcal{O}_1$, where the value \mathcal{O}_n of the option at maturity is nothing but $\mathcal{O}_n^Y(S_n)$. It is therefore clear that evaluating the price of an option with early exercise features means to simulate the evolution of the underlying asset price (to get the \mathcal{O}_i^Y) and to calculate a

(usually large) number of expectation conditional probabilities.

Standard Numerical Procedures

To value derivatives when analytical formulae are not available, appropriate numerical techniques have to be advocated. They involve the use of Monte Carlo (MC) simulation, binomial trees (and their improvements) and finite-difference methods [Hull (2000); Wilmott *et. al.* (1993)].

A natural way to simulate price paths is to discretize (6.37) as

$$\ln S(t + \Delta t) - \ln S(t) = A\Delta t + \sigma\epsilon\sqrt{\Delta t},$$

or, equivalently,

$$S(t + \Delta t) = S(t) \exp \left[A\Delta t + \sigma\epsilon\sqrt{\Delta t} \right], \quad (6.42)$$

which is correct for any $\Delta t > 0$, even if finite. Given the spot price S_0 , i.e., the price of the asset at time $t = 0$, one can extract from a standardized normal distribution a value ϵ_k , ($k = 1, \dots, n$) for the random variable ϵ to simulate one possible path followed by the price by means of (6.42):

$$S(k\Delta t) = S((k-1)\Delta t) \exp \left[A\Delta t + \sigma\epsilon_k\sqrt{\Delta t} \right].$$

Iterating the procedure m times, one can simulate m price paths $\{(S_0, S_1^{(j)}, S_2^{(j)}, \dots, S_n^{(j)} \equiv S_T^{(j)}) : j = 1, \dots, m\}$ and evaluate the price of the option. In such a MC simulation of the stochastic dynamics of asset price (Monte Carlo random walk) the mean values $E[\mathcal{O}_i | S_{i-1}]$, $i = 1, \dots, n$ are given by

$$E[\mathcal{O}_i | S_{i-1}] = \frac{\mathcal{O}_i^{(1)} + \mathcal{O}_i^{(2)} + \dots + \mathcal{O}_i^{(m)}}{m},$$

with no need to calculate transition probabilities because, through the extraction of the possible ϵ values, the paths are automatically weighted according to the probability distribution function of (6.39). Unfortunately, this method leads to an estimated value whose numerical error is proportional to $m^{-1/2}$. Thus, even if it is powerful because of the possibility to control the paths and to impose additional constraints (as it is usually required by exotic and path-dependent options), the MC random walk is extremely time consuming when precise predictions are required and appropriate variance reduction procedures have to be used to save CPU time [Hull (2000)]. This difficulty can be overcome by means of the method of the

binomial trees and its extensions (see [Hull (2000)] and references therein), whose main idea stands in a deterministic choice of the possible paths to limit the number of intermediate points. At each time step the price S_i is assumed to have only two choices: increase to the value $uS_i, u > 1$ or decrease to $dS_i, 0 < d < 1$, where the parameters u and d are given in terms of σ and Δt in such a way to give the correct values for the mean and variance of stock price changes over the time interval Δt . Also finite difference methods are known in the literature [Hull (2000)] as an alternative to time-consuming MC simulations. They give the value of the derivative by solving the differential equation satisfied by the derivative, by converting it into a difference equation. Although tree approaches and finite difference methods are known to be faster than the MC random walk, they are difficult to apply when a detailed control of the history of the derivative is required and are also computationally time consuming when a number of stochastic variables is involved [Hull (2000)]. It follows that the development of efficient and fast computational algorithms to price financial derivatives is still a key issue in financial analysis.

6.3.6.2 Option Pricing via Path Integrals

Recall that the path integral method is an integral formulation of the dynamics of a stochastic process. It is a suitable framework for the calculation of the transition probabilities associated to a given stochastic process, which is seen as the convolution of an infinite sequence of infinitesimal short-time steps [Bennati *et al.* (1999)]. For the problem of option pricing, the path-integral method can be employed for the explicit calculation of the expectation values of the quantities of financial interest, given by integrals of the form [Bennati *et al.* (1999)]

$$E[\mathcal{O}_i|S_{i-1}] = \int dz_i p(z_i|z_{i-1}) \mathcal{O}_i(e^{z_i}), \quad (6.43)$$

where $z = \ln S$ and $p(z_i|z_{i-1})$ is the transition probability. $E[\mathcal{O}_i|S_{i-1}]$ is the conditional expectation value of some functional \mathcal{O}_i of the stochastic process. For example, for an European call option at the maturity T the quantity of interest will be $\max\{S_T - X, 0\}$, X being the strike price. As already emphasized, and discussed in the literature [Hull (2000); Wilmott *et al.* (1993); Potters *et al.* (2001); Rosa-Clot and Taddei (2002); Matacz (2002)], the computational complexity associated to this calculation is generally great: in the case of exotic options, with path-dependent

and early exercise features, integrals of the type (6.43) cannot be analytically solved. As a consequence, we demand two things from a path integral framework: a very quick way to estimate the transition probability associated to a stochastic process (6.37) and a clever choice of the integration points with which evaluate the integrals (6.43). In particular, our aim is to develop an efficient calculation of the probability distribution without losing information on the path followed by the asset price during its time evolution.

Transition Probability

The probability distribution function related to a SDE verifies the *Chapman–Kolmogorov equation* [Paul and Baschnagel (1999)]

$$p(z''|z') = \int dz p(z''|z)p(z|z'), \quad (6.44)$$

which states that the probability (density) of a transition from the value z' (at time t') to the value z'' (at time t'') is the ‘summation’ over all the possible intermediate values z of the probability of separate and consequent transitions $z' \rightarrow z, z \rightarrow z''$. As a consequence, if we consider a finite time interval $[t', t'']$ and we apply a time slicing, by considering $n+1$ subintervals of length $\Delta t = (t'' - t')/n + 1$, we can write, by iteration of (6.44)

$$p(z''|z') = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} dz_1 \cdots dz_n p(z''|z_n) p(z_n|z_{n-1}) \cdots p(z_1|z'),$$

which, thanks to (6.38), can be written as [Montagna *et. al.* (2002)]

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} dz_1 \cdots dz_n \frac{1}{\sqrt{(2\pi\sigma^2\Delta t)^{n+1}}} \exp \left\{ -\frac{1}{2\sigma^2\Delta t} \sum_{k=1}^{n+1} [z_k - (z_{k-1} + A\Delta t)]^2 \right\}. \quad (6.45)$$

In the limit $n \rightarrow \infty, \Delta t \rightarrow 0$ such that $(n+1)\Delta t = (t'' - t')$ (infinite sequence of infinitesimal time steps), the expression (6.45), as explicitly shown in [Bennati *et. al.* (1999)], exhibits a Lagrangian structure and it is possible to express the transition probability in the path integral formalism as a convolution of the form [Bennati *et. al.* (1999)]

$$p(z'', t''|z', t') = \int_C \mathcal{D}[\sigma^{-1}\tilde{z}] \exp \left\{ -\int_{t'}^{t''} L(\tilde{z}(\tau), \dot{\tilde{z}}(\tau); \tau) d\tau \right\},$$

where L is the Lagrangian, given by

$$L(\tilde{z}(\tau), \dot{\tilde{z}}(\tau); \tau) = \frac{1}{2\sigma^2} [\dot{\tilde{z}}(\tau) - A]^2,$$

and the integral is performed (with functional measure $\mathcal{D}[\cdot]$) over the paths $\tilde{z}(\cdot)$ belonging to \mathcal{C} , i.e., all the continuous functions with constraints $\tilde{z}(t') \equiv z'$, $\tilde{z}(t'') \equiv z''$. As carefully discussed in [Bennati *et al.* (1999)], a path integral is well defined only if both a continuous formal expression and a discretization rule are given. As done in many applications, the Itô prescription is adopted here (see subsection 6.2.5 above).

A first, naïve evaluation of the transition probability (6.45) can be performed via Monte Carlo simulation, by writing (6.45) as

$$p(z'', t'' | z', t') = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \prod_i^n dg_i \frac{1}{\sqrt{2\pi\sigma^2\Delta t}} \exp \left\{ -\frac{1}{2\sigma^2\Delta t} [z'' - (z_n + A\Delta t)]^2 \right\}, \quad (6.46)$$

in terms of the variables g_i defined by the relation

$$dg_k = \frac{dz_k}{\sqrt{2\pi\sigma^2\Delta t}} \exp \left\{ -\frac{1}{2\sigma^2\Delta t} [z_k - (z_{k-1} + A\Delta t)]^2 \right\}, \quad (6.47)$$

and extracting each g_i from a gaussian distribution of mean $z_{k-1} + A\Delta t$ and variance $\sigma^2\Delta t$. However, as we will see, this method requires a large number of calls to get a good precision. This is due to the fact that each g_i is related to the previous g_{i-1} , so that this implementation of the path integral approach can be seen to be equivalent to a naïve MC simulation of random walks, with no variance reduction.

By means of appropriate manipulations [Schulman (1981)] of the integrand entering (6.45), it is possible, as shown in the following, to get a path integral expression which will contain a factorized integral with a constant kernel and a consequent variance reduction. If we define $z'' = z_{n+1}$ and $y_k = z_k - kA\Delta t$, $k = 1, \dots, n$, we can express the *transition probability distribution* as

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} dy_1 \cdots dy_n \frac{1}{\sqrt{(2\pi\sigma^2\Delta t)^{n+1}}} \exp \left\{ -\frac{1}{2\sigma^2\Delta t} \sum_{k=1}^{n+1} [y_k - y_{k-1}]^2 \right\}, \quad (6.48)$$

in order to get rid of the contribution of the drift parameter. Now let us

extract from the argument of the exponential function a quadratic form

$$\begin{aligned} \sum_{k=1}^{n+1} [y_k - y_{k-1}]^2 &= y_0^2 - 2y_1y_0 + y_1^2 + y_1^2 - 2y_1y_2 + \dots + y_{n+1}^2 \\ &= y^t M y + [y_0^2 - 2y_1y_0 + y_{n+1}^2 - 2y_ny_{n+1}], \end{aligned} \quad (6.49)$$

by introducing the nD array y and the $n \times n$ matrix M defined as [Montagna *et. al.* (2002)]

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{pmatrix}, \quad M = \begin{pmatrix} 2 & -1 & 0 & \dots & \dots & 0 \\ -1 & 2 & -1 & 0 & \dots & 0 \\ 0 & -1 & 2 & -1 & \dots & 0 \\ 0 & \dots & -1 & 2 & -1 & 0 \\ 0 & \dots & \dots & -1 & 2 & -1 \\ 0 & \dots & \dots & \dots & -1 & 2 \end{pmatrix}, \quad (6.50)$$

where M is a real, symmetric, non singular and tridiagonal matrix. In terms of the eigenvalues m_i of the matrix M , the contribution in (6.49) can be written as

$$y^t M y = w^t O^t M O w = w^t M_d w = \sum_{i=1}^n m_i w_i^2, \quad (6.51)$$

by introducing the orthogonal matrix O which diagonalizes M , with $w_i = O_{ij} y_j$. Because of the orthogonality of O , the Jacobian

$$J = \det \left| \frac{dw_i}{dy_k} \right| = \det |O_{ki}|,$$

of the transformation $y_k \rightarrow w_k$ equals 1, so that $\prod_{i=1}^n dw_i = \prod_{i=1}^n dy_i$. After some algebra, (6.49) can be written as

$$\begin{aligned} \sum_{k=1}^{n+1} [y_k - y_{k-1}]^2 &= \sum_{i=1}^n m_i w_i^2 + y_0^2 - 2y_1y_0 + y_{n+1}^2 - 2y_ny_{n+1} = \\ &= \sum_{i=1}^n m_i \left[w_i - \frac{(y_0 O_{1i} + y_{n+1} O_{ni})}{m_i} \right]^2 + y_0^2 + y_{n+1}^2 - \sum_{i=1}^n \frac{(y_0 O_{1i} + y_{n+1} O_{ni})^2}{m_i}. \end{aligned} \quad (6.52)$$

Now, if we introduce new variables h_i obeying the relation

$$dh_i = \sqrt{\frac{m_i}{2\pi\sigma^2\Delta t}} \exp \left\{ -\frac{m_i}{2\sigma^2\Delta t} \left[w_i - \frac{(y_0 O_{1i} + y_{n+1} O_{ni})}{m_i} \right]^2 \right\} dw_i, \quad (6.53)$$

it is possible to express the *finite-time probability distribution* $p(z''|z')$ as [Montagna *et. al.* (2002)]

$$\begin{aligned} & \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \prod_{i=1}^n dy_i \frac{1}{\sqrt{(2\pi\sigma^2\Delta t)^{n+1}}} \exp \left\{ -\frac{1}{2\sigma^2\Delta t} \sum_{k=1}^{n+1} [y_k - y_{k-1}]^2 \right\} \\ &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \prod_{i=1}^n dw_i \frac{1}{\sqrt{(2\pi\sigma^2\Delta t)^{n+1}}} e^{-(y_0^2 + y_{n+1}^2)/2\sigma^2\Delta t} \\ & \quad \times \exp \left\{ -\frac{1}{2\sigma^2\Delta t} \sum_{i=1}^n \left[m_i \left(w_i - \frac{(y_0 O_{1i} + y_{n+1} O_{ni})}{m_i} \right)^2 - \frac{(y_0 O_{1i} + y_{n+1} O_{ni})^2}{m_i} \right] \right\} \\ &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \prod_{i=1}^n dh_i \frac{1}{\sqrt{2\pi\sigma^2\Delta t \det(M)}} \\ & \quad \times \exp \left\{ -\frac{1}{2\sigma^2\Delta t} \left[y_0^2 + y_{n+1}^2 + \sum_{i=1}^n \frac{(y_0 O_{1i} + y_{n+1} O_{ni})^2}{m_i} \right] \right\}. \end{aligned} \quad (6.54)$$

The probability distribution function, as given by (6.54), is an integral whose kernel is a constant function (with respect to the integration variables) and which can be factorized into the n integrals

$$\int_{-\infty}^{+\infty} dh_i \exp \left\{ -\frac{1}{2\sigma^2\Delta t} \frac{(y_0 O_{1i} + y_{n+1} O_{ni})^2}{m_i} \right\}, \quad (6.55)$$

given in terms of the h_i , which are gaussian variables that can be extracted from a normal distribution with mean $(y_0 O_{1i} + y_{n+1} O_{ni})^2/m_i$ and variance $\sigma^2\Delta t/m_i$. Differently to the first, naïve implementation of the path integral, now each h_i is no longer dependent on the previous h_{i-1} , and importance sampling over the paths is automatically accounted for.

It is worth noticing that, by means of the extraction of the random variables h_i , we are creating price paths, since at each intermediate time t_i the asset price is given by

$$S_i = \exp \left\{ \sum_{k=1}^n O_{ik} h_k + iA\Delta t \right\}. \quad (6.56)$$

Therefore, this path integral algorithm can be easily adapted to the cases in which the derivative to be valued has, in the time interval $[0, T]$, additional

constraints, as in the case of interesting path-dependent options, such as Asian and barrier options [Hull (2000)].

Integration Points

The above illustrated method represents a powerful and fast tool to calculate the transition probability in the path integral framework and it can be employed if we need to value a generic option with maturity T and with possibility of anticipated exercise at times $t_i = i\Delta t$ ($n\Delta t = T$) [Montagna *et al.* (2002)]. As a consequence of this time slicing, one must numerically evaluate $n - 1$ mean values of the type (9), in order to check at any time t_i , and for any value of the stock price, whether early exercise is more convenient with respect to holding the option for a future time. To keep under control the computational complexity and the time of execution, it is mandatory to limit as far as possible the number of points for the integral evaluation. This means that we would like to have a linear growth of the number of integration points with the time. Let us suppose to evaluate each mean value

$$E[\mathcal{O}_i|S_{i-1}] = \int dz_i p(z_i|z_{i-1})\mathcal{O}_i(e^{z_i}),$$

with p integration points, i.e., considering only p fixed values for z_i . To this end, we can create a grid of possible prices, according to the dynamics of the stochastic process as given by (6.37)

$$z(t + \Delta t) - z(t) = \ln S(t + \Delta t) - \ln S(t) = A\Delta t + \epsilon\sigma\sqrt{\Delta t}. \quad (6.57)$$

Starting from z_0 , we thus evaluate the expectation value $E[\mathcal{O}_1|S_0]$ with $p = 2m + 1$, $m \in \mathbb{N}$ values of z_1 centered on the mean value $E[z_1] = z_0 + A\Delta t$ and which differ from each other of a quantity of the order of $\sigma\sqrt{\Delta t}$

$$z_1^j = z_0 + A\Delta t + j\sigma\sqrt{\Delta t}, \quad (j = -m, \dots, +m).$$

Going on like this, we can evaluate each expectation value $E[\mathcal{O}_2|z_1^j]$ get from each one of the z_1 's created above with p values for z_2 centered around the mean value

$$E[z_2|z_1^j] = z_1^j + A\Delta t = z_0 + 2A\Delta t + j\sigma\sqrt{\Delta t}.$$

Iterating the procedure until the maturity, we create a deterministic grid of points such that, at a given time t_i , there are $(p - 1)i + 1$ values of

z_i , in agreement with the request of linear growth. This procedure of selection of integration points, together with the calculation of the transition probability previously described, is the basis of the path integral simulation of the price of a generic option.

By applying the results derived above, we have at disposal an efficient path integral algorithm both for the calculation of transition probabilities and the evaluation of option prices. In [Montagna *et. al.* (2002)] the application of the above path–integral method to European and American options in the BSM model was illustrated and comparisons with the results were get with the standard procedures known in the literature were shown. First, the path integral simulation of the probability distribution of the logarithm of the stock prices, $p(\ln S)$, as a function of the logarithm of the stock price, for a BSM–like stochastic model, was given by (6.36). Once the transition probability has been computed, the price of an option could be computed in a path integral approach as the conditional expectation value of a given functional of the stochastic process. For example, the price of an European call option was given by

$$\mathcal{C} = e^{-r(T-t)} \int_{-\infty}^{+\infty} dz_f p(z_f, T|z_i, t) \max[e^{z_f} - X, 0], \quad (6.58)$$

while for an European put it will be

$$\mathcal{P} = e^{-r(T-t)} \int_{-\infty}^{+\infty} dz_f p(z_f, T|z_i, t) \max[X - e^{z_f}, 0], \quad (6.59)$$

where r is the risk–free interest rate. Therefore just 1D integrals need to be evaluated and they can be precisely computed with standard quadrature rules.

6.3.6.3 Continuum Limit and American Options

In the specific case of an *American option*, the possibility of exercise at any time up to the expiration date allows to develop, within the path integral formalism, a specific algorithm, which, as shown in the following, is precise and very quick [Montagna *et. al.* (2002)].

Given the time slicing considered above, the case of American options requires the limit $\Delta t \rightarrow 0$ which, putting $\sigma \rightarrow 0$, leads to a delta–like transition probability

$$p(z, t + \Delta t | z_t, t) \approx \delta(z - z_t - A\Delta t).$$

This means that, apart from volatility effects, the price z_i at time t_i will have a value remarkably close to the expected value $\bar{z} = z_{i-1} + A\Delta t$, given by the drift growth. In order to take care of the volatility effects, a possible solution is to estimate the integral of interest, i.e.,

$$E[\mathcal{O}_i|S_{i-1}] = \int_{-\infty}^{+\infty} dz p(z|z_{i-1}) \mathcal{O}_i(e^z), \quad (6.60)$$

by inserting in (6.60) the analytical expression for the $p(z|z_{i-1})$ transition probability

$$\begin{aligned} p(z|z_{i-1}) &= \frac{1}{\sqrt{2\pi\Delta t\sigma^2}} \exp\left\{-\frac{(z - z_{i-1} - A\Delta t)^2}{2\sigma^2\Delta t}\right\} \\ &= \frac{1}{\sqrt{2\pi\Delta t\sigma^2}} \exp\left\{-\frac{(z - \bar{z})^2}{2\sigma^2\Delta t}\right\}, \end{aligned}$$

together with a Taylor expansion of the kernel function $\mathcal{O}_i(e^z) = f(z)$ around the expected value \bar{z} . Hence, up to the second-order in $z - \bar{z}$, the kernel function becomes

$$f(z) = f(\bar{z}) + (z - \bar{z})f'(\bar{z}) + \frac{1}{2}f''(\bar{z})(z - \bar{z})^2 + O((z - \bar{z})^3),$$

which induces

$$E[\mathcal{O}_i|S_{i-1}] = f(\bar{z}) + \frac{\sigma^2}{2}f''(\bar{z}) + \dots,$$

since the first derivative does not give contribution to (6.60), being the integral of an odd function over the whole z range. The second derivative can be numerically estimated as

$$f''(\bar{z}) = \frac{1}{\delta_\sigma^2} [f(\bar{z} + \delta_\sigma) - 2f(\bar{z}) + f(\bar{z} - \delta_\sigma)],$$

with $\delta_\sigma = O(\sigma\sqrt{\Delta t})$, as dictated by the dynamics of the stochastic process.

6.3.7 Application: *Nonlinear Dynamics of Complex Nets*

Recall that many systems in nature, such as neural nets, food webs, metabolic systems, co-authorship of papers, the worldwide web, etc. can be represented as *complex networks*, or *small-world networks* (see, e.g., [Watts and Strogatz (1998); Dorogovtsev and Mendes (2003)]). In particular, it has been recognized that many networks have scale-free topology; the distribution of the degree obeys the power law, $P(k) \sim k^{-\gamma}$. The study

of the scale-free network now attracts the interests of many researchers in mathematics, physics, engineering and biology [Ichinomiya (2004)].

Another important aspect of complex networks is their dynamics, describing e.g., the spreading of viruses in the Internet, change of populations in a food web, and synchronization of neurons in a brain. In particular, [Ichinomiya (2004)] studied the synchronization of the random network of oscillators. His work follows the previous studies (see [Strogatz (2000)]) that showed that mean-field type synchronization, that Kuramoto observed in globally-coupled oscillators [Kuramoto (1984)], appeared also in the small-world networks.

6.3.7.1 Continuum Limit of the Kuramoto Net

Ichinomiya started with the standard network with N nodes, described by a variant of the *Kuramoto model*. Namely, at each node, there exists an oscillator and the phase of each oscillator θ_i is evolving according to

$$\dot{\theta}_i = \omega_i + K \sum_j a_{ij} \sin(\theta_j - \theta_i), \quad (6.61)$$

where K is the coupling constant, a_{ij} is 1 if the nodes i and j are connected, and 0 otherwise; ω_i is a random number, whose distribution is given by the function $N(\omega)$.

For the analytic study, it is convenient to use the *continuum limit equation*. We define $P(k)$ as the distribution of nodes with degree k , and $\rho(k, \omega; t, \theta)$ the density of oscillators with phase θ at time t , for given ω and k . We assume that $\rho(k, \omega; t, \theta)$ is normalized as

$$\int_0^{2\pi} \rho(k, \omega; t, \theta) d\theta = 1.$$

For simplicity, we also assume $N(\omega) = N(-\omega)$. Thus, we suppose that the collective oscillation corresponds to the stable solution, $\dot{\rho} = 0$.

Now we construct the continuum limit equation for the network of oscillators. The evolution of ρ is determined by the *continuity equation* $\partial_t \rho = -\partial_\theta(\rho v)$, where v is defined by the continuum limit of the r.h.s of (6.61). Because one randomly selected edge connects to the node of degree k , frequency ω , phase θ with the probability $kP(k)N(\omega)\rho(k, \omega; t, \theta) / \int dk k P(k)$,

$\rho(k, \omega; t, \theta)$ obeys the equation

$$\partial_t \rho(k, \omega; t, \theta) = -\partial_\theta [\rho(k, \omega; t, \theta) (\omega + \frac{Kk \int d\omega' \int dk' \int d\theta' N(\omega') P(k') k' \rho(k', \omega'; t, \theta') \sin(\theta - \theta')}{\int dk' P(k') k'})].$$

The mean-field solution of this equation was studied by [Ichinomiya (2004)].

6.3.7.2 Path-Integral Approach to Complex Nets

Recently, [Ichinomiya (2005)] introduced the *path-integral* (see subsection 4.4.6 above) *approach* in studying the dynamics of complex networks. He considered the stochastic generalization of the Kuramoto network (6.61), given by

$$\dot{x}_i = f_i(x_i) + \sum_{j=1}^N a_{ij} g(x_i, x_j) + \xi_i(t), \quad (6.62)$$

where $f_i = f_i(x_i)$ and $g_{ij} = g(x_i, x_j)$ are functions of network activations x_i , $\xi_i(t)$ is a *random force* that satisfies $\langle \xi_i(t) \rangle = 0$, $\langle \xi_i(t) \xi_j(t') \rangle = \delta_{ij} \delta(t-t') \sigma^2$. He assumed $x_i = x_{i,0}$ at $t = 0$. In order to discuss the dynamics of this system, he introduced the so-called *Matrin-Siggia-Rose* (MSR) *generating functional* Z given by [de Dominicis (1978)]

$$Z[\{l_{ik}\}, \{\bar{l}_{ik}\}] = \left(\frac{1}{\pi}\right)^{NN_t} \left\langle \int \prod_{i=1}^N \prod_{k=0}^{N_t} dx_{ik} d\bar{x}_{ik} e^{-S} \exp(l_{ik} x_{ik} + \bar{l}_{ik} \bar{x}_{ik})^J \right\rangle,$$

where the *action* S is given by

$$S = \sum_{ik} \left[\frac{\sigma^2 \Delta t}{2} \bar{x}_{ik}^2 + i \bar{x}_{ik} \{x_{ik} - x_{i,k-1} - \Delta t (f_i(x_{i,k-1}) + \sum_j a_{ij} g(x_{i,k-1}, x_{j,k-1}))\} \right],$$

and $\langle \cdots \rangle$ represents the average over the ensemble of networks. J is the functional Jacobian term,

$$J = \exp \left(-\frac{\Delta t}{2} \sum_{ijk} \frac{\partial (f_i(x_{ik}) + a_{ij} g(x_{ik}, x_{jk}))}{\partial x_{ik}} \right).$$

Ichinomiya considered such a form of the network model in which

$$a_{ij} = \begin{cases} 1 & \text{with probability } p_{ij}, \\ 0 & \text{with probability } 1 - p_{ij}. \end{cases}$$

Note that p_{ij} can be a function of variables such as i or j . For example, in the 1D chain model, p_{ij} is 1 if $|i - j| = 1$, else it is 0. The average over all networks can be expressed as

$$\left\langle \exp \left[\sum_{ik} i \Delta t \bar{x}_{ik} \sum_j a_{ij} g(x_{i,k-1}, x_{j,k-1}) \right] \right\rangle = \prod_{ij} \left[p_{ij} \exp \left\{ \sum_k i \Delta t \bar{x}_{ik} g(x_{i,k-1}, x_{j,k-1}) \right\} + 1 - p_{ij} \right],$$

so we get

$$\langle e^{-S} \rangle = \exp(-S_0) \prod_{ij} \left[p_{ij} \exp \left\{ \sum_k i \Delta t \bar{x}_{ik} g(x_{i,k-1}, x_{j,k-1}) \right\} + 1 - p_{ij} \right],$$

where $S_0 = \sum_{ik} \frac{\sigma^2 \Delta t}{2} \bar{x}_{ik}^2 + i \bar{x}_{ik} \{x_{ik} - x_{i,k-1} - \Delta t f_i(x_{i,k-1})\}.$

This expression can be applied to the dynamics of any complex network model. [Ichinomiya (2005)] applied this model to analysis of the *Kuramoto transition in random sparse networks*.

6.3.8 Application: Dissipative Quantum Brain Model

The *conservative brain* model was originally formulated within the framework of the quantum field theory (QFT) by [Ricciardi and Umezawa (1967)] and subsequently developed in [Stuart *et al.* (1978); Stuart *et al.* (1979); Jibu and Yasue (1995); Jibu *et al.* (1996)]. The conservative brain model has been recently extended to the *dissipative quantum dynamics* in the work of G. Vitiello and collaborators [Vitiello (1995); Alfinito and Vitiello (2000); Pessa and Vitiello (1999); Vitiello (2001); Pessa and Vitiello (2003); Pessa and Vitiello (2004)].

The canonical quantization procedure of a dissipative system requires to include in the formalism also the system representing the environment (usually the heat bath) in which the system is embedded. One possible way to do that is to depict the environment as the time-reversal image of the system [Celeghini *et al.* (1992)]: the environment is thus described as the *double* of the system in the time-reversed dynamics (the system image in the mirror of time).

Within the framework of dissipative QFT, the brain system is described in terms of an *infinite collection of damped harmonic oscillators* A_κ (the

simplest prototype of a dissipative system) representing the DWQ [Vitiello (1995)]. Now, the collection of damped harmonic oscillators is ruled by the Hamiltonian [Vitiello (1995); Celeghini *et al.* (1992)]

$$H = H_0 + H_I, \quad \text{with} \\ H_0 = \hbar\Omega_\kappa(A_\kappa^\dagger A_\kappa - \tilde{A}_\kappa^\dagger \tilde{A}_\kappa), \quad H_I = i\hbar\Gamma_\kappa(A_\kappa^\dagger \tilde{A}_\kappa^\dagger - A_\kappa \tilde{A}_\kappa),$$

where Ω_κ is the frequency and Γ_κ is the damping constant. The \tilde{A}_κ modes are the ‘time-reversed mirror image’ (i.e., the ‘mirror modes’) of the A_κ modes. They are the doubled modes, representing the environment modes, in such a way that κ generically labels their degrees of freedom. In particular, we consider the damped harmonic oscillator (DHO)

$$m\ddot{x} + \gamma\dot{x} + \kappa x = 0, \quad (6.63)$$

as a simple prototype for dissipative systems (with intention that thus get results also apply to more general systems). The damped oscillator (6.63) is a non-Hamiltonian system and therefore the customary canonical quantization procedure cannot be followed. However, one can face the problem by resorting to well known tools such as the *density matrix* ρ and the *Wigner function* W .

Let us start with the special case of a *conservative particle* in the absence of friction γ , with the standard Hamiltonian, $H = -(\hbar\partial_x)^2/2m + V(x)$.

Recall (from the previous subsection) that the *density matrix equation of motion*, i.e., *quantum Liouville equation*, is given by

$$i\hbar\dot{\rho} = [H, \rho]. \quad (6.64)$$

The density matrix function ρ is defined by

$$\langle x + \frac{1}{2}y | \rho(t) | x - \frac{1}{2}y \rangle = \psi^*(x + \frac{1}{2}y, t) \psi(x - \frac{1}{2}y, t) \equiv W(x, y, t),$$

with the associated standard expression for the *Wigner function* (see [Feynman and Hibbs (1965)]),

$$W(p, x, t) = \frac{1}{2\pi\hbar} \int W(x, y, t) e^{(-i\frac{py}{\hbar})} dy.$$

Now, in the coordinate x -representation, by introducing the notation

$$x_\pm = x \pm \frac{1}{2}y, \quad (6.65)$$

the Liouville equation (6.64) can be expanded as

$$\begin{aligned} i\hbar \partial_t \langle x_+ | \rho(t) | x_- \rangle = \\ \left\{ -\frac{\hbar^2}{2m} [\partial_{x_+}^2 - \partial_{x_-}^2] + [V(x_+) - V(x_-)] \right\} \langle x_+ | \rho(t) | x_- \rangle, \end{aligned} \quad (6.66)$$

while the Wigner function $W(p, x, t)$ is now given by

$$\begin{aligned} i\hbar \partial_t W(x, y, t) &= H_o W(x, y, t), \quad \text{with} \\ H_o &= \frac{1}{m} p_x p_y + V(x + \frac{1}{2}y) - V(x - \frac{1}{2}y), \\ \text{and} \quad p_x &= -i\hbar \partial_x, \quad p_y = -i\hbar \partial_y. \end{aligned} \quad (6.67)$$

The new Hamiltonian H_o (6.67) may be get from the corresponding Lagrangian

$$L_o = m\dot{x}\dot{y} - V(x + \frac{1}{2}y) + V(x - \frac{1}{2}y). \quad (6.68)$$

In this way, Vitiello concluded that the density matrix and the Wigner function formalism *required*, even in the conservative case (with zero mechanical resistance γ), the introduction of a ‘doubled’ set of coordinates, x_{\pm} , or, alternatively, x and y . One may understand this as related to the introduction of the ‘couple’ of indices *necessary* to label the density matrix elements (6.66).

Let us now consider the case of the *particle interacting* with a *thermal bath* at temperature T . Let f denote the *random force* on the particle at the position x due to the bath. The interaction Hamiltonian between the bath and the particle is written as

$$H_{int} = -fx. \quad (6.69)$$

Now, in the *Feynman–Vernon formalism* (see [Feynman (1972)]), the *effective action* $A[x, y]$ for the particle is given by

$$A[x, y] = \int_{t_i}^{t_f} L_o(\dot{x}, \dot{y}, x, y) dt + I[x, y],$$

with L_o defined by (6.68) and

$$e^{\frac{i}{\hbar} I[x, y]} = \langle (e^{-\frac{i}{\hbar} \int_{t_i}^{t_f} f(t)x_-(t)dt})_- (e^{\frac{i}{\hbar} \int_{t_i}^{t_f} f(t)x_+(t)dt})_+ \rangle, \quad (6.70)$$

where the symbol $\langle \cdot \rangle$ denotes the average with respect to the thermal bath; ‘ $(\cdot)_+$ ’ and ‘ $(\cdot)_-$ ’ denote time ordering and anti-time ordering, respectively; the coordinates x_{\pm} are defined as in (6.65). If the interaction between the

bath and the coordinate x (6.69) were turned off, then the operator f of the bath would develop in time according to $f(t) = e^{iH_\gamma t/\hbar} f e^{-iH_\gamma t/\hbar}$, where H_γ is the Hamiltonian of the isolated bath (decoupled from the coordinate x). $f(t)$ is then the force operator of the bath to be used in (6.70).

The interaction $I[x, y]$ between the bath and the particle has been evaluated in [Srivastava *et al.* (1995)] for a linear passive damping due to thermal bath by following Feynman–Vernon and Schwinger [Feynman and Hibbs (1965)]. The final result from [Srivastava *et al.* (1995)] is:

$$I[x, y] = \frac{1}{2} \int_{t_i}^{t_f} dt [x(t) F_y^{ret}(t) + y(t) F_x^{adv}(t)] \\ + \frac{i}{2\hbar} \int_{t_i}^{t_f} \int_{t_i}^{t_f} dt ds N(t-s) y(t) y(s),$$

where the retarded force on y , F_y^{ret} , and the advanced force on x , F_x^{adv} , are given in terms of the retarded and advanced Green functions $G_{ret}(t-s)$ and $G_{adv}(t-s)$ by

$$F_y^{ret}(t) = \int_{t_i}^{t_f} ds G_{ret}(t-s) y(s), \quad F_x^{adv}(t) = \int_{t_i}^{t_f} ds G_{adv}(t-s) x(s),$$

respectively. In (6.71), $N(t-s)$ is the *quantum noise* in the fluctuating random force given by: $N(t-s) = \frac{1}{2} \langle f(t) f(s) + f(s) f(t) \rangle$.

The real and the imaginary part of the action are given respectively by

$$\text{Re} [A[x, y]] = \int_{t_i}^{t_f} L dt, \quad (6.71)$$

$$L = m\dot{x}\dot{y} - \left[V(x + \frac{1}{2}y) - V(x - \frac{1}{2}y) \right] + \frac{1}{2} [x F_y^{ret} + y F_x^{adv}], \quad (6.72)$$

$$\text{and} \quad \text{Im} [A[x, y]] = \frac{1}{2\hbar} \int_{t_i}^{t_f} \int_{t_i}^{t_f} N(t-s) y(t) y(s) dt ds. \quad (6.73)$$

Equations (6.71–6.73), are *exact* results for linear passive damping due to the bath. They show that in the classical limit ‘ $\hbar \rightarrow 0$ ’ nonzero y yields an ‘unlikely process’ in view of the large imaginary part of the action implicit in (6.73). Nonzero y , indeed, may lead to a negative real exponent in the evolution operator, which in the limit $\hbar \rightarrow 0$ may produce a negligible contribution to the probability amplitude. On the contrary, at quantum level nonzero y accounts for quantum noise effects in the fluctuating random

force in the system–environment coupling arising from the imaginary part of the action (see [Srivastava *et al.* (1995)]).

When in (6.72) we use

$$F_y^{ret} = \gamma \dot{y} \quad \text{and} \quad F_x^{adv} = -\gamma \dot{x} \quad \text{we get,}$$

$$L(\dot{x}, \dot{y}, x, y) = m\dot{x}\dot{y} - V\left(x + \frac{1}{2}y\right) + V\left(x - \frac{1}{2}y\right) + \frac{\gamma}{2}(x\dot{y} - y\dot{x}). \quad (6.74)$$

By using

$$V\left(x \pm \frac{1}{2}y\right) = \frac{1}{2}\kappa(x \pm \frac{1}{2}y)^2$$

in (6.74), the DHO equation (6.63) and its complementary equation for the y coordinate

$$m\ddot{y} - \gamma\dot{y} + \kappa y = 0. \quad (6.75)$$

are derived. The y –oscillator is the time–reversed image of the x –oscillator (6.63). From the manifolds of solutions to equations (6.63) and (6.75), we could choose those for which the y coordinate is constrained to be zero, they simplify to

$$m\ddot{x} + \gamma\dot{x} + \kappa x = 0, \quad y = 0.$$

Thus we get the classical damped oscillator equation from a Lagrangian theory at the expense of introducing an ‘extra’ coordinate y , later constrained to vanish. Note that the constraint $y(t) = 0$ is *not* in violation of the equations of motion since it is a true solution to (6.63) and (6.75).

6.3.9 Application: *Cerebellum as a Neural Path–Integral*

Recall that human motion is naturally driven by synergistic action of more than 600 skeletal muscles. While the muscles generate driving torques in the moving joints, subcortical neural system performs both local and global (loco)motion control: first reflexly controlling contractions of individual muscles, and then orchestrating all the muscles into synergetic actions in order to produce efficient movements. While the local reflex control of individual muscles is performed on the *spinal control level*, the global integration of all the muscles into coordinated movements is performed within the *cerebellum*.

All hierarchical subcortical neuro-muscular physiology, from the bottom level of a single muscle fiber, to the top level of cerebellar muscular synergy, acts as a *temporal* $\langle out|in \rangle$ *reaction*, in such a way that the higher level acts as a command/control space for the lower level, itself representing an abstract image of the lower one:

- (1) At the *muscular level*, we have *excitation-contraction dynamics* [Hatze (1977a); Hatze (1978); Hatze (1977b)], in which $\langle out|in \rangle$ is given by the following sequence of nonlinear diffusion processes: *neural-action-potential* \rightsquigarrow *synaptic-potential* \rightsquigarrow *muscular-action-potential* \rightsquigarrow *excitation-contraction-coupling* \rightsquigarrow *muscle-tension-generating* [Ivancevic (1991); Ivancevic and Ivancevic (2006)]. Its purpose is the generation of muscular forces, to be transferred into driving torques within the joint anatomical geometry.
- (2) At the *spinal level*, $\langle out|in \rangle$ is given by *autogenetic-reflex stimulus-response control* [Houk (1979)]. Here we have a neural image of all individual muscles. The main purpose of the spinal control level is to give both positive and negative feedbacks to stabilize generated muscular forces within the ‘homeostatic’ (or, more appropriately, ‘homeokinetic’) limits. The individual muscular actions are combined into flexor-extensor (or agonist-antagonist) pairs, mutually controlling each other. This is the mechanism of *reciprocal innervation of agonists and inhibition of antagonists*. It has a purely mechanical purpose to form the so-called *equivalent muscular actuators* (EMAs), which would generate driving torques $T_i(t)$ for all movable joints.
- (3) At the *cerebellar level*, $\langle out|in \rangle$ is given by *sensory-motor integration* [Houk *et al.* (1996)]. Here we have an abstracted image of all autogenetic reflexes. The main purpose of the cerebellar control level is integration and fine tuning of the action of all active EMAs into a synchronized movement, by *supervising* the individual autogenetic reflex circuits. At the same time, to be able to perform in new and unknown conditions, the cerebellum is continuously adapting its own neural circuitry by unsupervised (self-organizing) learning. Its action is subconscious and automatic, both in humans and in animals.

Naturally, we can ask the question: Can we assign a single $\langle out|in \rangle$ measure to all these neuro-muscular stimulus-response reactions? We think that we can do it; so in this Letter, we propose the concept of *adaptive sensory-motor transition amplitude* as a unique measure for this temporal

$< out|in >$ relation. Conceptually, this $< out|in >$ –*amplitude* can be formulated as the ‘*neural path integral*’:

$$< out|in > \equiv \langle motor|sensory \rangle_{amplitude} = \int \mathcal{D}[w, x] e^{iS[x]}. \quad (6.76)$$

Here, the integral is taken over all *activated* (or, ‘fired’) *neural pathways* $x^i = x^i(t)$ of the cerebellum, connecting its input *sensory*–state with its output *motor*–state, symbolically described by *adaptive neural measure* $\mathcal{D}[w, x]$, defined by the weighted product (of discrete time steps)

$$\mathcal{D}[w, x] = \lim_{n \rightarrow \infty} \prod_{t=1}^n w^i(t) dx^i(t),$$

in which the *synaptic weights* $w^i = w^i(t)$, included in all active neural pathways $x^i = x^i(t)$, are updated by the unsupervised *Hebbian-like learning* rule 6.7, namely

$$w^i(t+1) = w^i(t) + \frac{\sigma}{\eta} (w_d^i(t) - w_a^i(t)), \quad (6.77)$$

where $\sigma = \sigma(t)$, $\eta = \eta(t)$ represent local neural *signal* and *noise* amplitudes, respectively, while superscripts *d* and *a* denote *desired* and *achieved* neural states, respectively. Theoretically, equations (6.76–6.77) define an ∞ –*dimensional neural network*. Practically, in a computer simulation we can use $10^7 \leq n \leq 10^8$, roughly corresponding to the number of neurons in the cerebellum.

The exponent term $S[x]$ in equation (6.76) represents the *autogenetic-reflex action*, describing reflexly-induced motion of all active EMAs, from their initial *stimulus*–state to their final *response*–state, along the family of extremal (i.e., Euler–Lagrangian) paths $x_{\min}^i(t)$. ($S[x]$ is properly derived in (6.80–6.81) below.)

6.3.9.1 Spinal Autogenetic Reflex Control

Recall (from Introduction) that at the spinal control level we have the autogenetic reflex *motor servo* [Houk (1979)], providing the local, reflex feedback loops for individual muscular contractions. A voluntary contraction force F of human skeletal muscle is reflexly excited (positive feedback $+F^{-1}$) by the responses of its *spindle receptors* to stretch and is reflexly inhibited (negative feedback $-F^{-1}$) by the responses of its *Golgi tendon organs*

to contraction. Stretch and unloading reflexes are mediated by combined actions of several autogenetic neural pathways, forming the *motor servo*.

In other words, branches of the afferent fibers also synapse with interneurons that inhibit motor neurons controlling the antagonistic muscles – *reciprocal inhibition*. Consequently, the stretch stimulus causes the antagonists to relax so that they cannot resist the shortening of the stretched muscle caused by the main reflex arc. Similarly, firing of the Golgi tendon receptors causes inhibition of the muscle contracting too strong and simultaneous *reciprocal activation* of its antagonist. Both mechanisms of reciprocal inhibition and activation performed by the autogenetic circuits $+F^{-1}$ and $-F^{-1}$, serve to generate the well-tuned EMA-driving torques T_i .

Now, once we have properly defined the symplectic musculo-skeletal dynamics [Ivancevic (2004)] on the biomechanical (momentum) phase-space manifold T^*M^N , we can proceed in formalizing its hierarchical subcortical neural control. By introducing the *coupling Hamiltonians* $H^m = H^m(q, p)$, selectively corresponding *only* to the $M \leq N$ *active joints*, we define the *affine Hamiltonian control function* $H_{aff} : T^*M^N \rightarrow \mathbb{R}$, in local canonical coordinates on T^*M^N given by (adapted from [Nijmeijer and van der Schaft (1990)] for the biomechanical purpose)

$$H_{aff}(q, p) = H_0(q, p) - H^m(q, p) T_m, \quad (m = 1, \dots, M \leq N), \quad (6.78)$$

where $T_m = T_m(t, q, p)$ are affine feedback torque one-forms, different from the initial driving torques T_i acting in all the joints. Using the affine Hamiltonian function (6.78), we get the *affine Hamiltonian servo-system* [Ivancevic (2004)],

$$\begin{aligned} \dot{q}^i &= \frac{\partial H_0(q, p)}{\partial p_i} - \frac{\partial H^m(q, p)}{\partial p_i} T_m, \\ \dot{p}_i &= -\frac{\partial H_0(q, p)}{\partial q^i} + \frac{\partial H^m(q, p)}{\partial q^i} T_m, \\ q^i(0) &= q_0^i, \quad p_i(0) = p_i^0, \quad (i = 1, \dots, N; \quad m = 1, \dots, M \leq N). \end{aligned} \quad (6.79)$$

The affine Hamiltonian control system (6.79) gives our formal description for the autogenetic spinal motor-servo for all $M \leq N$ activated (i.e., working) EMAs.

6.3.9.2 *Cerebellum – the Comparator*

Having, thus, defined the spinal reflex control level, we proceed to model the top subcortical commander/controller, the *cerebellum*. It is a brain region anatomically located at the bottom rear of the head (the hindbrain), directly above the brainstem, which is important for a number of subconscious and automatic motor functions, including motor learning. It processes information received from the motor cortex, as well as from proprioceptors and visual and equilibrium pathways, and gives ‘instructions’ to the motor cortex and other subcortical motor centers (like the basal nuclei), which result in proper balance and posture, as well as smooth, coordinated skeletal movements, like walking, running, jumping, driving, typing, playing the piano, etc. Patients with cerebellar dysfunction have problems with precise movements, such as walking and balance, and hand and arm movements. The cerebellum looks *similar in all animals*, from fish to mice to humans. This has been taken as evidence that it performs a common function, such as regulating motor learning and the timing of movements, in all animals. Studies of simple forms of motor learning in the vestibulo-ocular reflex and eye-blink conditioning are demonstrating that timing and amplitude of learned movements are encoded by the cerebellum.

The cerebellum is responsible for coordinating precisely timed $< out|in >$ activity by integrating motor output with ongoing sensory feedback. It receives extensive projections from sensory-motor areas of the cortex and the periphery and directs it back to premotor and motor cortex [Ghez (1990); Ghez (1991)]. This suggests a role in sensory-motor integration and the timing and execution of human movements. The cerebellum stores patterns of motor control for frequently performed movements, and therefore, its circuits are changed by experience and training. It was termed the *adjustable pattern generator* in the work of J. Houk and collaborators [Houk *et al.* (1996)]. Also, it has become the inspiring ‘brain-model’ in the recent robotic research [Schaal and Atkeson (1998); Schaal (1998)].

Comparing the number of its neurons ($10^7 - 10^8$), to the size of conventional neural networks, suggests that artificial neural nets *cannot* satisfactorily model the function of this sophisticated ‘super-bio-computer’, as its dimensionality is virtually infinite. Despite a lot of research dedicated to its structure and function (see [Houk *et al.* (1996)] and references there cited), the real nature of the cerebellum still remains a ‘mystery’.

The main function of the cerebellum as a motor controller is depicted

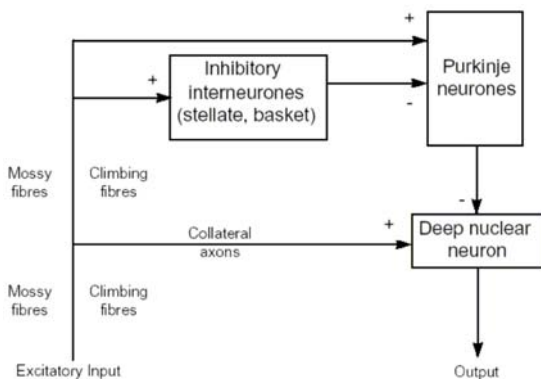


Fig. 6.5 Schematic *< out|in >* organization of the primary cerebellar circuit. In essence, excitatory inputs, conveyed by collateral axons of Mossy and Climbing fibers activate directly neurones in the Deep cerebellar nuclei. The activity of these latter is also modulated by the inhibitory action of the cerebellar cortex, mediated by the Purkinje cells.

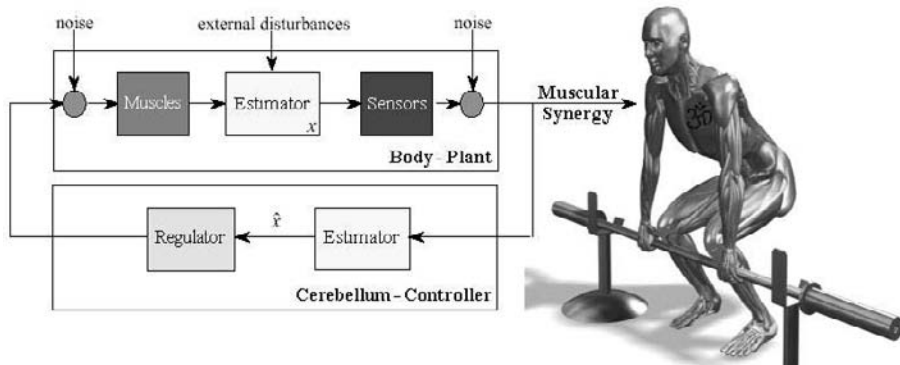


Fig. 6.6 The cerebellum as a motor controller.

in Figure 6.6. A coordinated movement is easy to recognize, but we know little about how it is achieved. In search of the neural basis of coordination, a model of spinocerebellar interactions was recently presented in [Apps and Garwicz (2005)], in which the structural and functional organizing principle is a division of the cerebellum into discrete micro-complexes. Each micro-complex is the recipient of a specific motor error signal - that is, a signal that conveys information about an inappropriate movement. These signals are encoded by spinal reflex circuits and conveyed to the cerebellar cortex

through climbing fibre afferents. This organization reveals salient features of cerebellar information processing, but also highlights the importance of systems level analysis for a fuller understanding of the neural mechanisms that underlie behavior.

6.3.9.3 Hamiltonian Action and Neural Path Integral

Here, we propose a *quantum-like adaptive control* approach to modelling the ‘cerebellar mystery’. Corresponding to the affine Hamiltonian control function (6.78) we define the *affine Hamiltonian control action*,

$$S_{aff}[q, p] = \int_{t_{in}}^{t_{out}} d\tau [p_i \dot{q}^i - H_{aff}(q, p)]. \quad (6.80)$$

From the affine Hamiltonian action (6.80) we further derive the associated expression for the *neural phase-space path integral* (in normal units), representing the *cerebellar sensory-motor amplitude* $\langle out|in \rangle$,

$$\begin{aligned} \langle q_{out}^i, p_i^{out} | q_{in}^i, p_i^{in} \rangle &= \int \mathcal{D}[w, q, p] e^{i S_{aff}[q, p]} \\ &= \int \mathcal{D}[w, q, p] \exp \left\{ i \int_{t_{in}}^{t_{out}} d\tau [p_i \dot{q}^i - H_{aff}(q, p)] \right\}, \\ \text{with} \quad \int \mathcal{D}[w, q, p] &= \int \prod_{\tau=1}^n \frac{w^i(\tau) dp_i(\tau) dq^i(\tau)}{2\pi}, \end{aligned} \quad (6.81)$$

where $w_i = w_i(t)$ denote the cerebellar synaptic weights positioned along its neural pathways, being continuously updated using the Hebbian-like self-organizing learning rule (6.77). Given the transition amplitude $\langle out|in \rangle$ (6.81), the *cerebellar sensory-motor transition probability* is defined as its absolute square, $|\langle out|in \rangle|^2$.

In (6.81), $q_{in}^i = q_{in}^i(t)$, $q_{out}^i = q_{out}^i(t)$; $p_i^{in} = p_i^{in}(t)$, $p_i^{out} = p_i^{out}(t)$; $t_{in} \leq t \leq t_{out}$, for all discrete time steps, $t = 1, \dots, n \rightarrow \infty$, and we are allowing for the affine Hamiltonian $H_{aff}(q, p)$ to depend upon all the ($M \leq N$) EMA-angles and angular momenta collectively. Here, we actually systematically took a discretized differential time limit of the form $t_\sigma - t_{\sigma-1} \equiv d\tau$ (both σ and τ denote discrete time steps) and wrote $\frac{(q_\sigma^i - q_{\sigma-1}^i)}{(t_\sigma - t_{\sigma-1})} \equiv \dot{q}^i$. For technical details regarding the path integral calculations on Riemannian and symplectic manifolds (including the standard regularization procedures), see [Klauder (1997); Klauder (2000)].

Now, motor learning occurring in the cerebellum can be observed using functional MR imaging, showing changes in the cerebellar action potential, related to the motor tasks (see, e.g., [Mascalchi *et. al.* (2002)]). To account for these electro-physiological currents, we need to add the *source* term $J_i(t)q^i(t)$ to the affine Hamiltonian action (6.80), (the current $J_i = J_i(t)$ acts as a source $J_i A^i$ of the *cerebellar electrical potential* $A^i = A^i(t)$),

$$S_{aff}[q, p, J] = \int_{t_{in}}^{t_{out}} d\tau [p_i \dot{q}^i - H_{aff}(q, p) + J_i q^i],$$

which, subsequently gives the cerebellar path integral with the action potential source, coming either from the motor cortex or from other subcortical areas.

Note that the standard *Wick rotation*: $t \mapsto it$ (see [Klauder (1997); Klauder (2000)]), makes all our path integrals real, i.e.,

$$\int \mathcal{D}[w, q, p] e^{i S_{aff}[q, p]} \xrightarrow{\text{Wick}} \int \mathcal{D}[w, q, p] e^{-S_{aff}[q, p]},$$

while their subsequent discretization gives the standard thermodynamic *partition functions*,

$$Z = \sum_j e^{-w_j E^j / T}, \quad (6.82)$$

where E^j is the energy eigenvalue corresponding to the affine Hamiltonian $H_{aff}(q, p)$, T is the temperature-like environmental control parameter, and the sum runs over all energy eigenstates (labelled by the index j). From (6.82), we can further calculate all statistical and thermodynamic system properties (see [Feynman (1972)]), as for example, *transition entropy* $S = k_B \ln Z$, etc.

6.3.10 Path Integrals via Jets: Perturbative Quantum Fields

Recall that an elegant way to make geometrical path integrals rigorous is to formulate them using the jet formalism. In this way the covariant Hamiltonian field systems were presented in [Bashkirov and Sardanashvily (2004)]. In this subsection we give a brief review of this perturbative quantum field model.

Let us quantize a Lagrangian system with the Lagrangian L_N (5.316) on the constraint manifold N_L (5.314). In the framework of a perturbative

quantum field theory, we should assume that $X = \mathbb{R}^n$ and $Y \rightarrow X$ is a trivial affine bundle. It follows that both the original coordinates $(x^\alpha, y^i, p_i^\alpha)$ and the adapted coordinates $(x^\alpha, y^i, p_a, p_A)$ on the Legendre bundle Π are global. Passing to field theory on an Euclidean space \mathbb{R}^n , we also assume that the matrix a in the Lagrangian L (5.312) is positive-definite, i.e., $a_{AA} > 0$.

Let us start from a Lagrangian (5.316) without gauge symmetries. Since the Lagrangian constraint space N_L can be equipped with the adapted coordinates p_A , the generating functional of Euclidean Green functions of the Lagrangian system in question reads [Bashkirov and Sardanashvily (2004)]

$$Z = \mathcal{N}^{-1} \int \exp\left\{\int (\mathcal{L}_N + \frac{1}{2} \text{Tr}(\ln \bar{\sigma}_0) + iJ_i y^i + iJ^A p_A) \omega\right\} \prod_x [dp_A(x)][dy(x)], \quad (6.83)$$

where \mathcal{L}_N is given by the expression (5.322) and $\bar{\sigma}_0$ is the square matrix

$$\bar{\sigma}_0^{AB} = M^{-1\alpha A} M^{-1\mu B} \sigma_{0\alpha\mu}^{ij} = \delta^{AB} (a_{AA})^{-1}.$$

The generating functional (6.83) a Gaussian integral of variables $p_A(x)$. Its integration with respect to $p_A(x)$ under the condition $J^A = 0$ restarts the generating functional

$$Z = \mathcal{N}^{-1} \int \exp\left\{\int (\mathcal{L} + iJ_i y^i) \omega\right\} \prod_x [dy(x)], \quad (6.84)$$

of the original Lagrangian field system on Y with the Lagrangian (5.312). However, the generating functional (6.83) cannot be rewritten with respect to the original variables p_i^μ , unless a is a nondegenerate matrix function.

In order to overcome this difficulty, let us consider a Lagrangian system on the whole Legendre manifold Π with the Lagrangian L_Π (5.319). Since this Lagrangian is constant along the fibres of the vector bundle $\Pi \rightarrow N_L$, an integration of the generating functional of this field model with respect to variables $p_a(x)$ should be finite. One can choose the generating functional in the form [Bashkirov and Sardanashvily (2004)]

$$Z = \mathcal{N}^{-1} \int \exp\left\{\int (\mathcal{L}_\Pi - \frac{1}{2} \sigma_{1\alpha\mu}^{ij} p_i^\alpha p_j^\mu + \frac{1}{2} \text{Tr}(\ln \sigma) + iJ_i y^i + iJ_\mu^i p_i^\mu) \omega\right\} \prod_x [dp(x)][dy(x)]. \quad (6.85)$$

Its integration with respect to momenta $p_i^\alpha(x)$ restarts the generating functional (6.84) of the original Lagrangian system on Y . In order to get the

generating functional (6.85), one can follow a procedure of quantization of gauge-invariant Lagrangian systems. In the case of the Lagrangian L_Π (5.319), this procedure is rather trivial, since the space of momenta variables $p_a(x)$ coincides with the translation subgroup of the gauge group $\text{Aut Ker } \sigma_0$.

Now let us suppose that the Lagrangian L_N (5.316) and, consequently, the Lagrangian L_Π (5.319) are invariant under some gauge group G_X of vertical automorphisms of the fibre bundle $Y \rightarrow X$ (and the induced automorphisms of $\Pi \rightarrow X$) which acts freely on the space of sections of $Y \rightarrow X$. Its infinitesimal generators are represented by vertical vector-fields $u = u^i(x^\mu, y^j)\partial_i$ on $Y \rightarrow X$ which induce the vector-fields

$$\bar{u} = u^i\partial_i - \partial_j u^i p_i^\alpha \partial_\alpha^j + d_\alpha u^i \partial_i^\alpha, \quad d_\alpha = \partial_\alpha + y_\alpha^i \partial_i, \quad (6.86)$$

on $\Pi \times J^1(X, Y)$. Let us also assume that G_X is indexed by m parameter functions $\xi^r(x)$ such that $u = u^i(x^\alpha, y^j, \xi^r)\partial_i$, where

$$u^i(x^\alpha, y^j, \xi^r) = u_r^i(x^\alpha, y^j)\xi^r + u_r^{i\mu}(x^\alpha, y^j)\partial_\mu \xi^r \quad (6.87)$$

are linear first-order differential operators on the space of parameters $\xi^r(x)$. The vector-fields $u(\xi^r)$ must satisfy the commutation relations

$$[u(\xi^q), u(\xi'^p)] = u(c_{pq}^r \xi'^p \xi^q),$$

where c_{pq}^r are structure constants. The Lagrangian L_Π (5.319) is invariant under the above gauge transformations iff its Lie derivative $\mathfrak{L}_{\bar{u}}L_\Pi$ along vector-fields (6.86) vanishes, i.e.,

$$(u^i\partial_i - \partial_j u^i p_i^\alpha \partial_\alpha^j + d_\alpha u^i \partial_i^\alpha)\mathcal{L}_\Pi = 0. \quad (6.88)$$

Since the operator $\mathfrak{L}_{\bar{u}}$ is linear in momenta p_i^μ , the condition (6.88) falls into the independent conditions

$$(u^k\partial_k - \partial_j u^k p_k^\nu \partial_\nu^j + d_\nu u^j \partial_j^\nu)(p_i^\alpha \mathcal{F}_\alpha^i) = 0, \quad (6.89)$$

$$(u^k\partial_k - \partial_j u^k p_k^\nu \partial_\nu^j)(\sigma_{0\lambda\mu}^{ij} p_i^\alpha p_j^\mu) = 0, \quad (6.90)$$

$$u^i\partial_i c' = 0. \quad (6.91)$$

It follows that the Lagrangian L_Π is gauge-invariant iff its three summands are separately gauge-invariant.

Note that, if the Lagrangian L_Π on Π is gauge-invariant, the original Lagrangian L (5.312) is also invariant under the same gauge transforma-

tions. Indeed, one gets at once from the condition (6.89) that

$$\bar{u}(\mathcal{F}_\mu^i) = \partial_j u^i \mathcal{F}_\mu^j, \quad (6.92)$$

i.e., the quantity \mathcal{F} is transformed as the dual of momenta p . Then the condition (6.90) shows that the quantity $\sigma_0 p$ is transformed by the same law as \mathcal{F} . It follows that the term $a\mathcal{F}\mathcal{F}$ in the Lagrangian L (5.312) is transformed exactly as $a(\sigma_0 p)(\sigma_0 p) = \sigma_0 p p$, i.e., is gauge-invariant. Then this Lagrangian is gauge-invariant due to the equality (6.91).

Since $\mathcal{S}_\alpha^i = y_\alpha^i - \mathcal{F}_\alpha^i$, one can derive from the formula (6.92) the transformation law of \mathcal{S} ,

$$\bar{u}(\mathcal{S}_\mu^i) = d_\mu u^i - \partial_j u^i \mathcal{F}_\mu^j = d_\mu u^i - \partial_j u^i (y_\mu^j - \mathcal{S}_\mu^j) = \partial_\mu u^i + \partial_j u^i \mathcal{S}_\mu^j. \quad (6.93)$$

This expression shows that the gauge group G_X acts freely on the space of sections $\mathcal{S}(x)$ of the fibre bundle $\widehat{\text{Ker}} \hat{L} \rightarrow Y$ in the splitting (5.309). Let the number m of parameters of the gauge group G_X do not exceed the fibre dimension of $\widehat{\text{Ker}} \hat{L} \rightarrow Y$. Then some combinations $b^{r\mu}_i \mathcal{S}_\mu^i$ of \mathcal{S}_μ^i can be used as the gauge condition

$$b^{r\mu}_i \mathcal{S}_\mu^i(x) - \alpha^r(x) = 0,$$

similar to the generalized Lorentz gauge in *Yang–Mills gauge theory*.

Now we turn to path-integral quantization of a Lagrangian system with the gauge-invariant Lagrangian L_Π (5.319). In accordance with the well-known quantization procedure, let us modify the generating functional (6.85) as follows [Bashkirov and Sardanashvily (2004)]

$$\begin{aligned} Z &= \mathcal{N}^{-1} \int \exp \left\{ \int (\mathcal{L}_\Pi - \frac{1}{2} \sigma_{1\alpha\mu}^{ij} p_i^\alpha p_j^\mu \right. \\ &\quad \left. + \frac{1}{2} \text{Tr}(\ln \sigma) - \frac{1}{2} h_{rs} \alpha^r \alpha^s + i J_i y^i + i J_\mu^\mu p_i^\mu) \omega \right\} \\ &\quad \Delta \prod_x \times^r \delta(b^{r\mu}_i \mathcal{S}_\mu^i(x) - \alpha^r(x)) [d\alpha(x)] [dp(x)] [dy(x)] \\ &= \mathcal{N}'^{-1} \int \exp \left\{ \int (\mathcal{L}_\Pi - \frac{1}{2} \sigma_{1\alpha\mu}^{ij} p_i^\alpha p_j^\mu + \frac{1}{2} \text{Tr}(\ln \sigma) \right. \\ &\quad \left. - \frac{1}{2} h_{rs} b^{r\mu}_i b^{s\alpha}_j \mathcal{S}_\mu^i \mathcal{S}_\alpha^j + i J_i y^i + i J_\mu^\mu p_i^\mu) \omega \right\} \Delta \prod_x [dp(x)] [dy(x)], \\ &\quad \text{where} \quad \int \exp \left\{ \int \left(-\frac{1}{2} h_{rs} \alpha^r \alpha^s \right) \omega \right\} \prod_x [d\alpha(x)] \end{aligned} \quad (6.94)$$

is a Gaussian integral, while the factor Δ is defined by the condition

$$\Delta \int \prod_x \times^r \delta(u(\xi)(b^{r\mu}_i \mathcal{S}^i_\mu)) [d\xi(x)] = 1.$$

We have the linear second-order differential operator

$$M^r_s \xi^s = u(\xi)(b^{r\mu}_i \mathcal{S}^i_\mu(x)) = b^{r\mu}_i (\partial_\mu u^i(\xi) + \partial_j u^i(\xi) \mathcal{S}^j_\mu) \quad (6.95)$$

on the parameter functions $\xi(x)$, and get $\Delta = \det M$. Then the generating functional (6.95) takes the form

$$\begin{aligned} Z = \mathcal{N}'^{-1} \int \exp \{ & \int (\mathcal{L}_\Pi - \frac{1}{2} \sigma^{ij}_{\alpha\mu} p^\alpha_i p^\mu_j + \frac{1}{2} \text{Tr} \ln \sigma - \frac{1}{2} h_{rs} b^{r\mu}_i b^{s\alpha}_j \mathcal{S}^i_\mu \mathcal{S}^j_\alpha \\ & - \bar{c}_r M^r_s c^s + i J_i y^i + i J^\mu_i p^\mu_i) \omega \} \prod_x [d\bar{c}][dc][dp(x)][dy(x)], \end{aligned} \quad (6.96)$$

where \bar{c}_r , c^s are odd ghost fields. Integrating Z (6.96) with respect to momenta under the condition $J^\mu_\mu = 0$, we come to the generating functional

$$Z = \mathcal{N}'^{-1} \int \exp \{ \int (\mathcal{L} - \frac{1}{2} h_{rs} b^{r\mu}_i b^{s\alpha}_j \mathcal{S}^i_\mu \mathcal{S}^j_\alpha - \bar{c}_r M^r_s c^s + i J_i y^i) \omega \} \prod_x [d\bar{c}][dc][dy(x)] \quad (6.97)$$

of the original field model on Y with the gauge-invariant Lagrangian L (5.312).

Note that the Lagrangian

$$\mathcal{L}' = \mathcal{L} - \frac{1}{2} h_{rs} b^{r\mu}_i b^{s\alpha}_j \mathcal{S}^i_\mu \mathcal{S}^j_\alpha - \bar{c}_r M^r_s c^s \quad (6.98)$$

fails to be gauge-invariant, but it admits the so-called *BRST symmetry*¹⁰ whose *odd operator* reads

$$\begin{aligned} \vartheta = & u^i(x^\mu, y^i, c^s) \partial_i + d_\alpha u^i(x^\mu, y^i, c^s) \partial^\alpha_i + \bar{v}_r(x^\mu, y^i, y^\mu_i) \frac{\partial}{\partial \bar{c}_r} \\ & + v^r(x^\mu, y^i, c^s) \frac{\partial}{\partial c^r} + d_\alpha v^r(x^\mu, y^i, c^s) \frac{\partial}{\partial c^r_\alpha} + d_\mu d_\alpha v^r(x^\mu, y^i, c^s) \frac{\partial}{\partial c^r_{\mu\lambda}}, \\ d_\alpha = & \partial_\alpha + y^i_\alpha \partial_i + y^i_{\lambda\mu} \partial^\mu_i + c^r_\alpha \frac{\partial}{\partial c^r} + c^r_{\lambda\mu} \frac{\partial}{\partial c^r_{\mu\lambda}}. \end{aligned} \quad (6.99)$$

¹⁰Recall that the BRST formalism is a method of implementing first class constraints. The letters BRST stand for Becchi, Rouet, Stora, and (independently) Tyutin who discovered this formalism. It is a rigorous method to deal with quantum theories with gauge invariance.

Its components $u^i(x^\mu, y^i, c^s)$ are given by the expression (6.87) where parameter functions $\xi^r(x)$ are replaced with the ghosts c^r . The components \bar{v}_r and v^r of the BRST operator ϑ can be derived from the condition

$$\vartheta(\mathcal{L}') = -h_{rs}M_q^r b^{s\alpha}_j \mathcal{S}^j_\alpha c^q - \bar{v}_r M_q^r c^q + \bar{c}_r \vartheta(\vartheta(b^{r\alpha}_j \mathcal{S}^j_\alpha)) = 0$$

of the BRST invariance of \mathcal{L}' . This condition falls into the two independent relations

$$\begin{aligned} h_{rs}M_q^r b^{s\alpha}_j \mathcal{S}^j_\alpha + \bar{v}_r M_q^r &= 0, \\ \vartheta(c^q)(\vartheta(c^p)(b^{r\alpha}_j \mathcal{S}^j_\alpha)) &= u(c^p)(u(c^q)(b^{r\alpha}_j \mathcal{S}^j_\alpha)) + u(v^r)(b^{r\alpha}_j \mathcal{S}^j_\alpha) \\ &= u\left(\frac{1}{2}c_{pq}^r c^p c^q + v^r\right)(b^{r\alpha}_j \mathcal{S}^j_\alpha) = 0. \end{aligned}$$

Hence, we get: $\bar{v}_r = -h_{rs}b^{s\alpha}_j \mathcal{S}^j_\alpha$, and $v^r = -\frac{1}{2}c_{pq}^r c^p c^q$.

6.4 Sum over Geometries and Topologies

Recall that the term *quantum gravity* (or *quantum geometrodynamics*, or *quantum geometry*), is usually understood as a consistent fundamental quantum description of gravitational space-time geometry whose classical limit is Einstein's general relativity. Among the possible ramifications of such a theory are a model for the structure of space-time near the Planck scale, a consistent calculational scheme to calculate gravitational effects at all energies, a description of quantum geometry near space-time singularities and a non-perturbative quantum description of 4D black holes. It might also help us in understanding cosmological issues about the beginning and end of the universe, i.e., the so-called 'big bang' and 'big crunch' (see e.g., [Penrose (1989); Penrose (1994); Penrose (1997)]).

From what we know about the quantum dynamics of other fundamental interactions it seems eminently plausible that also the gravitational excitations should at very short scales be governed by quantum laws. Now, conventional *perturbative path integral expansions of gravity*, as well as perturbative expansion in the string coupling in the case of unified approaches, both have difficulty in finding any direct or indirect evidence for quantum gravitational effects, be they experimental or observational, which could give a feedback for model building. The outstanding problems mentioned above require a non-perturbative treatment; it is not sufficient to know the first few terms of a perturbation series. The real goal is to search for a

non-perturbative definition of such a theory, where the initial input of any fixed ‘background metric’ is inessential (or even undesirable), and where ‘space–time’ is determined *dynamically*. Whether or not such an approach necessarily requires the inclusion of higher dimensions and fundamental supersymmetry is currently unknown (see [Ambjørn and Kristjansen (1993); Ambjørn and Loll (1998); Ambjørn *et. al.* (2000a); Ambjørn *et. al.* (2000b); Ambjørn *et. al.* (2001a); Ambjørn *et. al.* (2001b); Ambjørn *et. al.* (2001c); Dasgupta and Loll (2001)]).

Such a non-perturbative viewpoint is very much in line with how one proceeds in classical geometrodynamics, where a metric space–time $(M, g_{\mu\nu})$ (+ matter) emerges only as a *solution* to the familiar *Einstein equation*

$$G_{\mu\nu}[g] \equiv R_{\mu\nu}[g] - \frac{1}{2}g_{\mu\nu}R[g] = -8\pi T_{\mu\nu}[\Phi], \quad (6.100)$$

which define the classical dynamics of fields $\Phi = \Phi^{\mu\nu}$ on the space $\mathcal{M}(M)$, the space of all metrics $g = g_{\mu\nu}$ on a given smooth manifold M . The analogous question we want to address in the quantum theory is: Can we get ‘quantum space–time’ as a solution to a set of non-perturbative quantum equations of motion on a suitable quantum analogue of $\mathcal{M}(M)$ or rather, of the space of geometries, $\text{Geom}(M) = \mathcal{M}(M)/\text{Diff}(M)$?

Now, this is not a completely straightforward task. Whichever way we want to proceed non-perturbatively, if we give up the privileged role of a flat, Minkowskian background space–time on which the quantization is to take place, we also have to abandon the central role usually played by the Poincaré group, and with it most standard quantum field–theoretic tools for regularization and renormalization. If one works in a continuum metric formulation of gravity, the symmetry group of the *Einstein–Hilbert action* is instead the group $\text{Diff}(M)$ of diffeomorphisms on M , which in terms of local charts are the smooth invertible coordinate transformations $x^\mu \mapsto y^\mu(x^\mu)$.

In the following, we will describe a non-perturbative path integral approach to quantum gravity, defined on the *space of all geometries*, without distinguishing any background metric structure [Loll (2001)]. This is closely related in spirit with the canonical approach of loop quantum gravity [Rovelli (1998)] and its more recent incarnations using so-called *spin networks* (see, e.g., [Oriti (2001)]). ‘Non-perturbative’ here means in a covariant context that the path sum or integral will have to be performed explicitly, and not just evaluated around its stationary points, which can only be

achieved in an appropriate regularization. The method we will employ uses a discrete lattice regularization as an intermediate step in the construction of the quantum theory.

6.4.1 *Simplicial Quantum Geometry*

In this section we will explain how one may construct a theory of quantum gravity from a non-perturbative path integral, using the method of Lorentzian dynamical triangulations. The method is minimal in the sense of employing standard tools from quantum field theory and the theory of critical phenomena and adapting them to the case of *generally covariant systems*, without invoking any symmetries beyond those of the classical theory. At an intermediate stage of the construction, we use a regularization in terms of simplicial *Regge geometries*, that is, piecewise linear manifolds. In this approach, ‘computing the path integral’ amounts to a conceptually simple and geometrically transparent ‘counting of geometries’, with additional weight factors which are determined by the EH action. This is done first of all at a regularized level. Subsequently, one searches for interesting continuum limits of these discrete models which are possible candidates for theories of quantum gravity, a step that will always involve a renormalization. From the point of view of statistical mechanics, one may think of Lorentzian dynamical triangulations as a new class of statistical models of Lorentzian random surfaces in various dimensions, whose building blocks are flat simplices which carry a ‘time arrow’, and whose dynamics is entirely governed by their intrinsic geometric properties.

Before describing the details of the construction, it may be helpful to recall the path integral representation for a 1D non-relativistic particle (see previous section). The time evolution of the particle’s wave function ψ may be described by the integral equation (6.3) above, where the *propagator*, or the *Feynman kernel* G , is defined through a limiting procedure (6.4). The time interval $t'' - t'$ has been discretized into N steps of length $\epsilon = (t'' - t')/N$, and the r.h.s. of (6.4) represents an integral over all piecewise linear paths $x(t)$ of a ‘virtual’ particle propagating from x' to x'' , illustrated in Figure 6.4 above.

The prefactor A^{-N} is a normalization and L denotes the Lagrangian function of the particle. Knowing the propagator G is tantamount to having solved the quantum dynamics. This is the simplest instance of a path

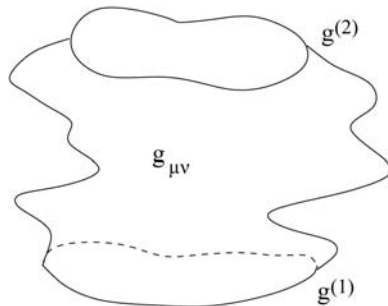


Fig. 6.7 The time-honored way of illustrating the gravitational path integral as the propagator from an initial to a final spatial boundary geometry (see text for explanation).

integral, and is often written schematically as

$$G(x', t'; x'', t'') = \int \mathcal{D}[x(t)] e^{iS[x(t)]}, \quad (6.101)$$

where $\mathcal{D}[x(t)]$ is a functional measure on the ‘space of all paths’, and the exponential weight depends on the classical action $S[x(t)]$ of a path. Recall also that this procedure can be defined in a mathematically clean way if we Wick-rotate the time variable t to imaginary values $t \mapsto \tau = it$, thereby making all integrals real [Reed and Simon (1975)].

Can a similar strategy work for the case of Einstein geometrodynamics? As an analogue of the particle’s position we can take the geometry $[g_{ij}(x)]$ (i.e., an equivalence class of spatial metrics) of a constant-time slice. Can one then define a gravitational propagator

$$G([g'_{ij}], [g''_{ij}]) = \int_{Geom(M)} \mathcal{D}[g_{\mu\nu}] e^{iS^{EH}[g_{\mu\nu}]} \quad (6.102)$$

from an initial geometry $[g']$ to a final geometry $[g'']$ (Figure 6.7) as a limit of some discrete construction analogous to that of the non-relativistic particle (6.4)? And crucially, what would be a suitable class of ‘paths’, that is, space-times $[g_{\mu\nu}]$ to sum over?

Now, to be able to perform the integration $\int \mathcal{D}[g_{\mu\nu}]$ in a meaningful way, the strategy we will be following starts from a regularized version of the space $Geom(M)$ of all geometries. A regularized path integral $G(a)$ can be defined which depends on an ultraviolet cutoff a and is *convergent* in a non-trivial region of the space of coupling constants. Taking the continuum limit corresponds to letting $a \rightarrow 0$. The resulting continuum theory – if it can be shown to exist – is then investigated with regard to its geometric

properties and in particular its semiclassical limit.

6.4.2 Discrete Gravitational Path Integrals

Trying to construct non-perturbative path integrals for gravity from sums over discretized geometries, using approach of *Lorentzian dynamical triangulations*, is not a new idea. Inspired by the successes of lattice gauge theory, attempts to describe quantum gravity by similar methods have been popular on and off since the late 70's. Initially the emphasis was on gauge-theoretic, first-order formulations of gravity, usually based on (compactified versions of) the Lorentz group, followed in the 80's by 'quantum Regge calculus', an attempt to represent the gravitational path integral as an integral over certain piecewise linear geometries (see [Williams (1997)] and references therein), which had first made an appearance in approximate descriptions of *classical* solutions of the Einstein equations. A variant of this approach by the name of 'dynamical triangulation(s)' attracted a lot of interest during the 90's, partly because it had proved a powerful tool in describing 2D quantum gravity (see the textbook [Ambjørn *et. al.* (1997)] and lecture notes [Ambjørn *et. al.* (2000a)] for more details).

The problem is that none of these attempts have so far come up with convincing evidence for the existence of an underlying continuum theory of 4D quantum gravity. This conclusion is drawn largely on the basis of numerical simulations, so it is by no means water-tight, although one can make an argument that the 'symptoms' of failure are related in the various approaches [Loll (1998)]. What goes wrong generically seems to be a dominance in the continuum limit of highly degenerate geometries, whose precise form depends on the approach chosen. One would expect that non-smooth geometries play a decisive role, in the same way as it can be shown in the particle case that the support of the measure in the continuum limit is on a set of nowhere differentiable paths. However, what seems to happen in the case of the path integral for 4-geometries is that the structures get *too* wild, in the sense of not generating, even at coarse-grained scales, an effective geometry whose dimension is anywhere near four.

The schematic phase diagram of Euclidean dynamical triangulations shown in Figure 6.8 gives an example of what can happen. The picture turns out to be essentially the same in both three and four dimensions: the model possesses infinite-volume limits everywhere along the critical line $k_3^{\text{crit}}(k_0)$, which fixes the bare cosmological constant as a function of the inverse Newton constant $k_0 \sim G_N^{-1}$. Along this line, there is a critical point

k_0^{crit} (which we now know to be of first-order in $d = 3, 4$) below which geometries generically have a very large effective or Hausdorff dimension.¹¹ Above k_0^{crit} we find the opposite phenomenon of ‘polymerization’: a typical element contributing to the state sum is a thin branched polymer, with one or more dimensions ‘curled up’ such that its effective dimension is around two.

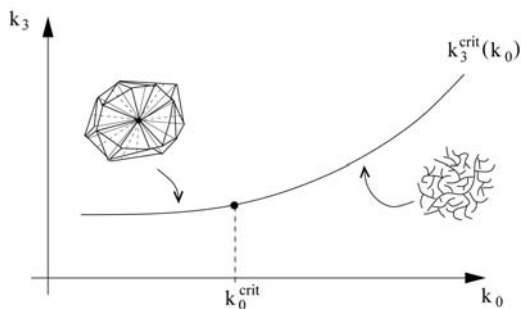


Fig. 6.8 The phase diagram of 3D and 4D Euclidean dynamical triangulations (see text for explanation).

This problem has to do with the fact that the gravitational action is unbounded below, causing potential havoc in Euclidean versions of the path integral. Namely, what all the above-mentioned approaches have in common is that they work from the outset with *Euclidean* geometries, and associated Boltzmann-type weights $\exp(-S^{\text{eu}})$ in the path integral. In other words, they integrate over ‘space-times’ which know nothing about time, light cones and causality. This is done mainly for technical reasons, since it is difficult to set up simulations with complex weights and since until recently a suitable Wick rotation was not known.

‘Lorentzian dynamical triangulations’, first proposed in [Ambjørn and Loll (1998)] and further elaborated in [Ambjørn *et. al.* (2000b); Ambjørn *et. al.* (2001a)] tries to establish a logical connection between the fact that non-perturbative path integrals were constructed for Euclidean instead of Lorentzian geometries and their apparent failure to lead to an interesting continuum theory.

¹¹In terms of geometry, this means that there are a few vertices at which the entire space-time ‘condenses’ in the sense that almost every other vertex in the simplicial space-time is about one link-distance away from them.

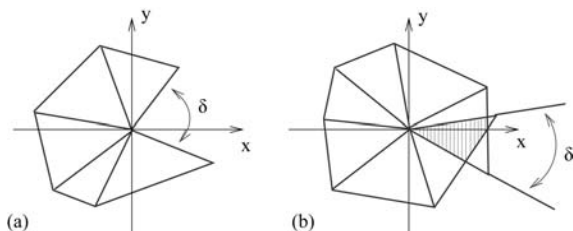


Fig. 6.9 Positive (a) and negative (b) space-like deficit angles δ (adapted from [Loll (2001); Loll (1998)]).

6.4.3 Regge Calculus

The use of simplicial methods in general relativity goes back to the pioneering work of Regge [Regge (1961)]. In classical applications one tries to approximate a classical space-time geometry by a triangulation, that is, a piecewise linear space get by gluing together flat simplicial building blocks, which in dimension d are dD generalizations of triangles. By ‘flat’ we mean that they are isometric to a subspace of dD Euclidean or Minkowski space. We will only be interested in gluings leading to genuine manifolds, which therefore look locally like an R^d . A nice feature of such simplicial manifolds is that their geometric properties are completely described by the discrete set $\{l_i^2\}$ of the squared lengths of their edges. Note that this amounts to a description of geometry *without the use of coordinates*. There is nothing to prevent us from re-introducing coordinate patches covering the piecewise linear manifold, for example, on each individual simplex, with suitable transition functions between patches. In such a coordinate system the metric tensor will then assume a definite form. However, for the purposes of formulating the path integral we will not be interested in doing this, but rather work with the edge lengths, which constitute a direct, regularized parametrization of the space $Geom(M)$ of geometries.

How precisely is the intrinsic geometry of a simplicial space, most importantly, its curvature, encoded in its edge lengths? A useful example to keep in mind is the case of dimension two, which can easily be visualized. A 2D piecewise linear space is a triangulation, and its scalar curvature $R(x)$ coincides with the *Gaussian curvature* (see section 3.10.1.3 above). One way of measuring this curvature is by parallel-transporting a vector around closed curves in the manifold. In our piecewise-flat manifold such a vector will always return to its original orientation *unless* it has surrounded lattice vertices v at which the surrounding angles did not add up to 2π , but

$\sum_{i \supset v} \alpha_i = 2\pi - \delta$, for $\delta \neq 0$, see Figure 6.9. The so-called deficit angle δ is precisely the rotation angle picked up by the vector and is a direct measure for the scalar curvature at the vertex. The operational description to get the scalar curvature in higher dimensions is very similar, one basically has to sum in each point over the Gaussian curvatures of all 2D submanifolds. This explains why in Regge calculus the curvature part of the EH action is given by a sum over building blocks of dimension $(d-2)$ which are the objects dual to those local 2D submanifolds. More precisely, the continuum curvature and volume terms of the action become

$$\frac{1}{2} \int_{\mathcal{R}} d^d x \sqrt{|\det g|} {}^{(d)}R \longrightarrow \sum_{i \in \mathcal{R}} \text{Vol}(i^{\text{th}} (d-2)\text{-simplex}) \delta_i \quad (6.103)$$

$$\int_{\mathcal{R}} d^d x \sqrt{|\det g|} \longrightarrow \sum_{i \in \mathcal{R}} \text{Vol}(i^{\text{th}} d\text{-simplex}) \quad (6.104)$$

in the simplicial discretization. It is then a simple exercise in trigonometry to express the volumes and angles appearing in these formulas as functions of the edge lengths l_i , both in the Euclidean and the Minkowskian case.

The approach of dynamical triangulations uses a certain class of such simplicial space-times as an explicit, regularized realization of the space $\text{Geom}(M)$. For a given volume N_d , this class consists of all gluings of manifold-type of a set of N_d simplicial building blocks of top-dimension d whose edge lengths are restricted to take either one or one out of two values. In the Euclidean case we set $l_i^2 = a^2$ for all i , and in the Lorentzian case we allow for both space- and time-like links with $l_i^2 \in \{-a^2, a^2\}$, where the geodesic distance a serves as a short-distance cutoff, which will be taken to zero later. Coming from the classical theory this may seem a grave restriction at first, but this is indeed not the case. Firstly, keep in mind that for the purposes of the quantum theory we want to sample the space of geometries ‘ergodically’ at a coarse-grained scale of order a . This should be contrasted with the classical theory where the objective is usually to approximate a given, *fixed* space-time to within a length scale a . In the latter case one typically requires a much finer topology on the space of metrics or geometries. It is also straightforward to see that no local curvature degrees of freedom are suppressed by fixing the edge lengths; deficit angles in all directions are still present, although they take on only a discretized set of values. In this sense, in dynamical triangulations all geometry is in the gluing of the fundamental building blocks. This is dual to how quantum Regge calculus is set up, where one usually fixes a triangulation T and then

‘scans’ the space of geometries by letting the l_i ’s run continuously over all values compatible with the triangular inequalities.

In a nutshell, Lorentzian dynamical triangulations give a definite meaning to the ‘integral over geometries’, namely, as a sum over inequivalent Lorentzian gluings T over any number N_d of d -simplices,

$$\int_{\text{Geom}(M)} \mathcal{D}[g_{\mu\nu}] e^{iS[g_{\mu\nu}]} \xrightarrow{\text{LDT}} \sum_{T \in \mathcal{T}} \frac{1}{C_T} e^{iS^{\text{Reg}}(T)}, \quad (6.105)$$

where the symmetry factor $C_T = |\text{Aut}(T)|$ on the r.h.s. is the order of the automorphism group of the triangulation, consisting of all maps of T onto itself which preserve the connectivity of the simplicial lattice. We will specify below what precise class \mathcal{T} of triangulations should appear in the summation.

It follows from the above that in this formulation all curvatures and volumes contributing to the *Regge simplicial action* come in discrete units. This can be illustrated by the case of a 2D triangulation with Euclidean signature, which according to the prescription of dynamical triangulations consists of *equilateral* triangles with squared edge lengths $+a^2$. All interior angles of such a triangle are equal to $\pi/3$, which implies that the deficit angle at any vertex v can take the values $2\pi - k_v\pi/3$, where k_v is the number of triangles meeting at v . As a consequence, the Einstein–Regge action S^{Reg} assumes the simple form

$$S^{\text{Reg}}(T) = \kappa_{d-2} N_{d-2} - \kappa_d N_d, \quad (6.106)$$

where the coupling constants $\kappa_i = \kappa_i(\lambda, G_N)$ are simple functions of the bare cosmological and Newton constants in d dimensions. Substituting this into the path sum in (6.105) leads to

$$Z(\kappa_{d-2}, \kappa_d) = \sum_{N_d} e^{-i\kappa_d N_d} \sum_{N_{d-2}} e^{i\kappa_{d-2} N_{d-2}} \sum_{T|_{N_d, N_{d-2}}} \frac{1}{C_T}, \quad (6.107)$$

The point of taking separate sums over the numbers of d - and $(d-2)$ -simplices in (6.107) is to make explicit that ‘doing the sum’ is tantamount to the combinatorial problem of *counting* triangulations of a given volume and number of simplices of codimension 2 (corresponding to the last summation in (6.107)).¹² It turns out that at least in two space–time dimensions the counting of geometries can be done completely explicitly,

¹²The symmetry factor C_T is almost always equal to 1 for large triangulations.

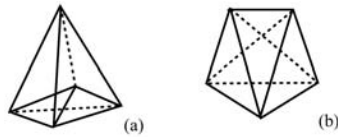


Fig. 6.10 Two types of Minkowskian 4-simplices in 4D (see text for explanation).

turning both Lorentzian and Euclidean quantum gravity into exactly soluble statistical models.

6.4.4 Lorentzian Path Integral

Now, the simplicial building blocks of the models are taken to be pieces of Minkowski space, and their edges have squared lengths $+a^2$ or $-a^2$. For example, the two types of 4-simplices that are used in Lorentzian dynamical triangulations in dimension four are shown in Figure 6.10. The first of them has four time-like and six space-like links (and therefore contains 4 time-like and 1 space-like tetrahedron), whereas the second one has six time-like and four space-like links (and contains 5 time-like tetrahedra). Since both are subspaces of flat space with signature $(-+++)$, they possess well-defined light-cone structures everywhere [Loll (2001); Loll (1998)].

In general, gluings between pairs of d -simplices are only possible when the metric properties of their $(d-1)$ -faces match. Having local light cones implies causal relations between pairs of points in local neighborhoods. Creating closed time-like curves will be avoided by requiring that all space-times contributing to the path sum possess a global ‘time’ function t . In terms of the triangulation this means that the d -simplices are arranged such that their space-like links all lie in slices of constant integer t , and their time-like links interpolate between adjacent spatial slices t and $t+1$. Moreover, with respect to this time, we will not allow for any *spatial* topology changes¹³.

This latter condition is always satisfied in classical applications, where ‘trouser points’ like the one depicted in Figure 6.11 are ruled out by the requirement of having a non-degenerate Lorentzian metric defined everywhere on M (it is geometrically obvious that the light cone and hence $g_{\mu\nu}$ must degenerate in at least one point along the ‘crotch’). Another way of thinking about such configurations (and their time-reversed counterparts)

¹³Note that if we were in the continuum and had introduced coordinates on space-time, such a statement would actually be diffeomorphism-invariant.

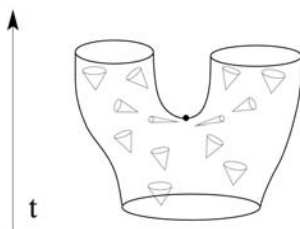


Fig. 6.11 At a branching point associated with a spatial topology change, light-cones get ‘squeezed’.

is that the causal past (future) of an observer changes discontinuously as her world-line passes near the singular point (see [Dowker (2002)] and references therein for related discussions about the issue of topology change in quantum gravity).

There is no *a priori* reason in the quantum theory to not relax some of these classical causality constraints. After all, as we stressed right at the outset, path integral histories are not in general classical solutions, nor can we attribute any other direct physical meaning to them individually. It might well be that one can construct models whose path integral configurations violate causality in this strict sense, but where this notion is somehow recovered in the resulting continuum theory. What the approach of Lorentzian dynamical triangulations has demonstrated is that *imposing causality constraints will in general lead to a different continuum theory*. This is in contrast with the intuition one may have that ‘including a few isolated singular points will not make any difference’. On the contrary, tampering with causality in this way is not innocent at all, as was already anticipated by Teitelboim many years ago [Teitelboim (1983)].

We want to point out that one cannot conclude from the above that spatial topology changes or even *fluctuations in the space-time topology* cannot be treated in the formulation of dynamical triangulations. However, if one insists on including geometries of variable topology in a Lorentzian discrete context, one has to come up with a prescription of how to weigh these singular points in the path integral, both before and after the Wick rotation [Dasgupta (2002)]. Maybe this can be done along the lines suggested in [Louko and Sorkin (1997)]; this is clearly an interesting issue for further research.

Having said this, we next have to address the question of the *Wick rotation*, in other words, of how to get rid of the factor of i in the exponent of

(6.107). Without it, this expression is an infinite sum (since the volume can become arbitrarily large) of complex terms whose convergence properties will be very difficult to establish. In this situation, a Wick rotation is simply a technical tool which – in the best of all worlds – enables us to perform the state sum and determine its continuum limit. The end result will have to be Wick-rotated back to Lorentzian signature.

Fortunately, Lorentzian dynamical triangulations come with a natural notion of Wick rotation, and the strategy we just outlined can be carried out explicitly in two space-time dimensions, leading to a unitary theory. In higher dimensions we do not yet have sufficient analytical control of the continuum theories to make specific statements about the *inverse* Wick rotation. Since we use the Wick rotation at an intermediate step, one can ask whether other Wick rotations would lead to the same result. Currently this is a somewhat academic question, since it is in practice difficult to find such alternatives. In fact, it is quite miraculous we have found a single prescription for Wick-rotating in our regularized setting, and it does not seem to have a direct continuum analogue (for more comments on this issue, see [Dasgupta and Loll (2001); Dasgupta (2002)]).

Our Wick rotation W in any dimension is an injective map from Lorentzian- to Euclidean-signature simplicial space-times. Using the notation T for a simplicial manifold together with length assignments l_s^2 and l_t^2 to its space- and time-like links, it is defined by

$$T^{\text{lor}} = (T, \{l_s^2 = a^2, l_t^2 = -a^2\}) \xrightarrow{W} T^{\text{eu}} = (T, \{l_s^2 = a^2, l_t^2 = a^2\}). \quad (6.108)$$

Note that we have not touched the connectivity of the simplicial manifold T , but only its metric properties, by mapping all time-like links of T into space-like ones, resulting in a Euclidean ‘space-time’ of equilateral building blocks. It can be shown [Ambjørn *et. al.* (2001a)] that at the level of the corresponding weight factors in the path integral this Wick rotation¹⁴ has precisely the desired effect of rotating to the exponentiated Regge action of the ‘Euclideanized’ geometry,

$$e^{iS(T^{\text{lor}})} \xrightarrow{W} e^{-S(T^{\text{eu}})}. \quad (6.109)$$

¹⁴To get a genuine Wick rotation and not just a discrete map, one introduces a complex parameter α in $l_t^2 = -\alpha a^2$. The proper prescription leading to (6.109) is then an analytic continuation of α from 1 to -1 through the lower-half complex plane.

The Euclideanized path sum after the Wick rotation has the form

$$\begin{aligned}
 Z^{\text{eu}}(\kappa_{d-2}, \kappa_d) &= \sum_T \frac{1}{C_T} e^{-\kappa_d N_d(T) + \kappa_{d-2} N_{d-2}(T)} \\
 &= \sum_{N_d} e^{-\kappa_d N_d} \sum_{T|N_d} \frac{1}{C_T} e^{\kappa_{d-2} N_{d-2}(T)} \\
 &= \sum_{N_d} e^{-\kappa_d N_d} e^{\kappa_d^{\text{crit}}(\kappa_{d-2}) N_d} \times \text{subleading}(N_d). \quad (6.110)
 \end{aligned}$$

In the last equality we have used that the number of Lorentzian triangulations of discrete volume N_d to leading order scales exponentially with N_d for large volumes. This can be shown explicitly in space–time dimension 2 and 3. For $d = 4$, there is strong (numerical) evidence for such an exponential bound for *Euclidean triangulations*, from which the desired result for the Lorentzian case follows (since W maps to a strict subset of all Euclidean simplicial manifolds).

From the functional form of the last line of (6.110) one can immediately read off some qualitative features of the phase diagram, an example of which appeared already earlier in Figure 6.8. Namely, the sum over geometries Z^{eu} converges for values $\kappa_d > \kappa_d^{\text{crit}}$ of the bare cosmological constant, and diverges (ie. is not defined) below this critical line. Generically, for all models of dynamical triangulations the infinite–volume limit is attained by approaching the critical line $\kappa_d^{\text{crit}}(\kappa_{d-2})$ from above, ie. from inside the region of convergence of Z^{eu} . In the process of taking $N_d \rightarrow \infty$ and the cutoff $a \rightarrow 0$, one gets a renormalized cosmological constant Λ through

$$(\kappa_d - \kappa_d^{\text{crit}}) = a^\mu \Lambda + O(a^{\mu+1}). \quad (6.111)$$

If the scaling is canonical (which means that the dimensionality of the renormalized coupling constant is the one expected from the classical theory), the exponent is given by $\mu = d$. Note that this construction requires a positive *bare* cosmological constant in order to make the state sum converge. Moreover, by virtue of relation (6.111) also the *renormalized* cosmological constant must be positive. Other than that, its numerical value is not determined by this argument, but by comparing observables of the theory which depend on Λ with actual physical measurements.¹⁵ Another interesting observation is that the inclusion of a sum over topologies in the discretized

¹⁵The non–negativity of the renormalized cosmological coupling may be taken as a first ‘prediction’ of our construction, which in the physical case of four dimensions is indeed in agreement with current observations.

sum (6.110) would lead to a super-exponential growth of at least $\propto N_d!$ of the number of triangulations with the volume N_d . Such a divergence of the path integral cannot be compensated by an additive renormalization of the cosmological constant of the kind outlined above.

There are ways in which one can sum divergent series of this type, for example, by performing a Borel sum. The problem with these stems from the fact that two different functions can share the same asymptotic expansion. Therefore, the series in itself is *not* sufficient to define the underlying theory uniquely. The non-uniqueness arises because of non-perturbative contributions to the path integral which are not represented in the perturbative expansion.¹⁶ In order to fix these uniquely, an independent, non-perturbative definition of the theory is necessary. Unfortunately, for dynamically triangulated models of quantum gravity, no such definitions have been found so far. In the context of 2D (Euclidean) quantum gravity this difficulty is known as the ‘absence of a physically motivated double-scaling limit’ [Ambjørn and Kristjansen (1993)].

Lastly, getting an interesting continuum limit may or may not require an additional fine-tuning of the inverse gravitational coupling κ_{d-2} , depending on the dimension d . In four dimensions, one would expect to find a second-order transition along the critical line, corresponding to local gravitonic excitations. The situation in $d = 3$ is less clear, but results get so far indicate that no fine-tuning of Newton’s constant is necessary [Ambjørn *et. al.* (2001b); Ambjørn *et. al.* (2001c)].

Before delving into the details, let us summarize briefly the results that have been get so far in the approach of Lorentzian dynamical triangulations. At the regularized level, that is, in the presence of a finite cutoff a for the edge lengths and an infrared cutoff for large space-time volume, they are well-defined statistical models of Lorentzian random geometries in $d = 2, 3, 4$. In particular, they obey a suitable notion of reflection-positivity and possess self-adjoint Hamiltonians.

The crucial questions are then to what extent the underlying combinatorial problems of counting all dD geometries with certain causal properties can be solved, whether continuum theories with non-trivial dynamics exist and how their bare coupling constants get renormalized in the process. What we know about Lorentzian dynamical triangulations so far is that they lead to continuum theories of quantum gravity in dimension 2 and 3. In $d = 2$, there is a complete analytic solution, which is distinct from the

¹⁶A field-theoretic example would be instantons and renormalons in QCD.

continuum theory produced by Euclidean dynamical triangulations. Also the matter-coupled model has been studied. In $d = 3$, there are numerical and partial analytical results which show that both a continuum theory exists and that it again differs from its Euclidean counterpart. Work on a more complete analytic solution which would give details about the geometric properties of the quantum theory is under way. In $d = 4$, the first numerical simulations are currently being set up. The challenge here is to do this for sufficiently large lattices, to be able to perform meaningful measurements. So far, we cannot make any statements about the existence and properties of a continuum theory in this physically most interesting case.

6.4.5 Application: *Topological Phase Transitions and Hamiltonian Chaos*

6.4.5.1 *Phase Transitions in Hamiltonian Systems*

Recall that *phase transitions* (PTs) are phenomena which bring about *qualitative* physical changes at the macroscopic level in presence of the same microscopic forces acting among the constituents of a system. Their mathematical description requires to translate into *quantitative* terms the mentioned qualitative changes. The standard way of doing this is to consider how the values of thermodynamic observables, get in laboratory experiments, vary with temperature, or volume, or an external field, and then to associate the experimentally observed discontinuities at a PT to the appearance of some kind of singularity entailing a loss of analyticity. Despite the smoothness of the statistical measures, after the *Yang–Lee Theorem* [Yang and Lee (1952)] we know that in the $N \rightarrow \infty$ limit non-analytic behaviors of thermodynamic functions are possible whenever the analyticity radius in the complex fugacity plane shrinks to zero, because this entails the loss of *uniform convergence* in N (number of degrees of freedom) of any sequence of real-valued thermodynamic functions, and all this depends on the distribution of the zeros of the grand canonical partition function. Also the other developments of the rigorous theory of PTs [Georgii (1988); Ruelle (1978)], identify PTs with the loss of analyticity.

In this subsection we will address a recently proposed geometric approach to thermodynamic phase transitions (see [Caiani *et al.* (1997); Franzosi *et al.* (1999); Franzosi *et al.* (2000); Franzosi and Pettini (2004)]). Given any Hamiltonian system, the configuration space can be equipped with a metric, in order to get a Riemannian geometrization of the dynam-

ics. At the beginning, several numerical and analytical studies of a variety of models showed that the fluctuation of the curvature becomes singular at the transition point. Then the following conjecture was proposed in [Cai-iani *et al.* (1997)]: The phase transition is determined by a change in the topology of the configuration space, and the loss of analyticity in the thermodynamic observables is nothing but a consequence of such topological change. The latter conjecture is also known as the *topological hypothesis*.

The topological hypothesis states that suitable topology changes of equipotential submanifolds of the Hamiltonian system's configuration manifold can entail thermodynamic phase transitions [Franzosi *et al.* (2000)]. The authors of the topological hypothesis gave both a theoretical argument and numerical demonstration in case of $2D$ lattice φ^4 model. They considered classical many-particle (or many-subsystem) systems described by standard *mechanical Hamiltonians*

$$H(p, q) = \sum_{i=1}^N \frac{p_i^2}{2m} + V(q), \quad (6.112)$$

where the coordinates $q^i = q^i(t)$ and momenta $p_i = p_i(t)$, ($i = 1, \dots, N$), have continuous values and the system's potential energy $V(q)$ is bounded below.

Now, assuming a large number of subsystems N , the statistical behavior of physical systems described by Hamiltonians of the type (6.112) is usually encompassed, in the system's canonical ensemble, by the *partition function* in the system's phase-space

$$\begin{aligned} Z_N(\beta) &= \int \prod_{i=1}^N dp_i dq^i e^{-\beta H(p, q)} = \left(\frac{\pi}{\beta}\right)^{\frac{N}{2}} \int \prod_{i=1}^N dq^i e^{-\beta V(q)} \\ &= \left(\frac{\pi}{\beta}\right)^{\frac{N}{2}} \int_0^\infty dv e^{-\beta v} \int_{M_v} \frac{d\sigma}{\|\nabla V\|}, \end{aligned} \quad (6.113)$$

where the last term is written using a *co-area formula* [Federer (1969)], and v labels the *equipotential hypersurfaces* M_v of the system's configuration manifold M ,

$$M_v = \{(q^1, \dots, q^N) \in \mathbb{R}^N | V(q^1, \dots, q^N) = v\}. \quad (6.114)$$

Equation (6.113) shows that for Hamiltonians (3.34) the relevant statistical

information is contained in the *canonical configurational partition function*

$$Z_N^C = \int \prod_{i=1}^N dq^i \exp[-\beta V(q)].$$

Therefore, partition function Z_N^C is decomposed – in the last term of equation (6.113) – into an infinite summation of geometric integrals, $\int_{M_v} d\sigma / \|\nabla V\|$, defined on the $\{M_v\}_{v \in \mathbb{R}}$. Once the microscopic interaction potential $V(q)$ is given, the configuration space of the system is automatically foliated into the family $\{M_v\}_{v \in \mathbb{R}}$ of these equipotential hypersurfaces. Now, from standard statistical mechanical arguments we know that, at any given value of the inverse temperature β , the larger the number N of particles the closer to $M_v \equiv M_{u_\beta}$ are the microstates that significantly contribute to the averages – computed through $Z_N(\beta)$ – of thermodynamic observables. The hypersurface M_{u_β} is the one associated with the average potential energy computed at a given β ,

$$u_\beta = (Z_N^C)^{-1} \int \prod_{i=1}^N dq^i V(q) \exp[-\beta V(q)].$$

Thus, at any β , if N is very large the effective support of the canonical measure shrinks very close to a single $M_v = M_{u_\beta}$.

Explicitly, the topological hypothesis reads: the basic origin of a phase transition lies in a suitable topology change of the $\{M_v\}$, occurring at some v_c . This topology change induces the singular behavior of the thermodynamic observables at a phase transition. By change of topology we mean that $\{M_v\}_{v < v_c}$ are not diffeomorphic to the $\{M_v\}_{v > v_c}$. In other words, canonical measure should ‘feel’ a big and sudden change of the topology of the equipotential hypersurfaces of its underlying support, the consequence being the appearance of the typical signals of a phase transition.

This point of view has the interesting consequence that – also at finite N – in principle *different* mathematical objects, i.e., manifolds of different cohomology type, could be associated to *different* thermodynamical phases, whereas from the point of view of measure theory [Yang and Lee (1952)] the only mathematical property available to signal the appearance of a phase transition is the loss of analyticity of the grand-canonical and canonical averages, a fact which is compatible with analytic statistical measures only in the mathematical $N \rightarrow \infty$ limit.

As it is conjectured that the counterpart of a phase transition is a breaking of diffeomorphicity among the surfaces M_v , it is appropriate to choose

a *diffeomorphism invariant* to probe if and how the topology of the M_v changes as a function of v . This is a very challenging task because we have to deal with high dimensional manifolds. Fortunately a topological invariant exists whose computation is feasible, yet demands a big effort. Recall (from subsection 3.10.1 above) that this is the *Euler characteristic*, a diffeomorphism invariant of the system's configuration manifold, expressing its fundamental topological information.

6.4.5.2 Geometry of the Largest Lyapunov Exponent

Now, the topological hypothesis has recently been promoted into a *topological Theorem* [Franzosi and Pettini (2004)]. The new Theorem says that non-analyticity is the 'shadow' of a more fundamental phenomenon occurring in the system's configuration manifold: a *topology change* within the family of equipotential hypersurfaces (6.114). This topological approach to PTs stems from the numerical study of the Hamiltonian dynamical counterpart of phase transitions, and precisely from the observation of *discontinuous* or *cuspy patterns*, displayed by the *largest Lyapunov exponent* at the transition energy (or temperature).

Recall that the *Lyapunov exponents* measure the strength of *dynamical chaos* and cannot be measured in laboratory experiments, at variance with thermodynamic observables, thus, being genuine dynamical observables they are only measurable in numerical simulations of the microscopic dynamics. To get a hold of the reason why the largest Lyapunov exponent λ_1 should probe configuration space topology, let us first remember that for standard Hamiltonian systems, λ_1 is computed by solving the *tangent dynamics equation* for Hamiltonian systems (see Jacobi equation of geodesic deviation (3.133)),

$$\ddot{\xi}_i + \left(\frac{\partial^2 V}{\partial q^i \partial q^j} \right)_{q(t)} \xi^j = 0, \quad (6.115)$$

which, for the nonlinear Hamiltonian system

$$\begin{aligned} \dot{q}^1 &= p_1, & \dot{p}_1 &= -\partial_{q^1} V, \\ &\dots & &\dots \\ \dot{q}^N &= p_N, & \dot{p}_N &= -\partial_{q^N} V, \end{aligned}$$

expands into *linearized Hamiltonian dynamics*

$$\begin{aligned}\dot{\xi}_1 &= \xi_{N+1}, & \dot{\xi}_{N+1} &= - \sum_{j=1}^N \left(\frac{\partial^2 V}{\partial q_1 \partial q_j} \right)_{q(t)} \xi_j, \\ &\dots & \dots & \\ \dot{\xi}_n &= \xi_{2N}, & \dot{\xi}_{2N} &= - \sum_{j=1}^N \left(\frac{\partial^2 V}{\partial q_N \partial q_j} \right)_{q(t)} \xi_j.\end{aligned}\tag{6.116}$$

Using (6.115) we can get the analytical expression for the largest Lyapunov exponent

$$\lambda_1 = \lim_{t \rightarrow \infty} \frac{1}{t} \log \frac{\left[\xi_1^2(t) + \dots + \xi_N^2(t) + \dot{\xi}_1^2(t) + \dots + \dot{\xi}_N^2(t) \right]^{1/2}}{\left[\xi_1^2(0) + \dots + \xi_N^2(0) + \dot{\xi}_1^2(0) + \dots + \dot{\xi}_N^2(0) \right]^{1/2}}.\tag{6.117}$$

If there are critical points of V in configuration space, that is points $q_c = [\bar{q}^1, \dots, \bar{q}^N]$ such that $\nabla V(q)|_{q=q_c} = 0$, according to the *Morse lemma* (see e.g., [Hirsch (1976)]), in the neighborhood of any critical point q_c there always exists a coordinate system $\bar{q}(t) = [\bar{q}^1(t), \dots, \bar{q}^N(t)]$ for which

$$V(\bar{q}) = V(q_c) - (\bar{q}^1)^2 - \dots - (\bar{q}^k)^2 + (\bar{q}^{k+1})^2 + \dots + (\bar{q}^N)^2,\tag{6.118}$$

where k is the index of the critical point, i.e., the number of negative eigenvalues of the Hessian of V . In the neighborhood of a critical point, equation (6.118) yields

$$\partial^2 V / \partial q^i \partial q^j = \pm \delta_{ij},$$

which, substituted into equation (6.115), gives k *unstable directions* which contribute to the exponential growth of the norm of the tangent vector $\xi = \xi(t)$. This means that the strength of dynamical chaos, measured by the largest Lyapunov exponent λ_1 , is affected by the existence of critical points of V . In particular, let us consider the possibility of a sudden variation, with the potential energy v , of the number of critical points (or of their indexes) in configuration space at some value v_c , it is then reasonable to expect that the pattern of $\lambda_1(v)$ – as well as that of $\lambda_1(E)$ since $v = v(E)$ – will be consequently affected, thus displaying *jumps* or *cusps* or other *singular patterns* at v_c .

On the other hand, recall that *Morse theory* teaches us that the existence of critical points of V is associated with topology changes of the hypersurfaces $\{M_v\}_{v \in \mathbb{R}}$, provided that V is a good *Morse function* (that is:

bounded below, with no vanishing eigenvalues of its Hessian matrix). Thus the existence of critical points of the potential V makes possible a conceptual link between dynamics and configuration space topology, which, on the basis of both direct and indirect evidence for a few particular models, has been formulated as a topological hypothesis about the relevance of topology for PTs phenomena (see [Franzosi *et al.* (2000); Franzosi and Pettini (2004); Grinza and Mossa (2004)]).

Here we give two simple examples of standard Hamiltonian systems of the form (6.112), namely Peyrard–Bishop system and mean-field XY model.

Peyrard–Bishop Hamiltonian System

The *Peyrard–Bishop system* [Peyrard and Bishop (1989)]¹⁷ exhibits a *second-order phase transition*. It is defined by the following potential energy

$$V(q) = \sum_{i=1}^N \left[\frac{K}{2} (q^{i+1} - q^i)^2 + D(e^{-aq^i} - 1)^2 + Dhaq^i \right], \quad (6.119)$$

which represents the energy of a string of N base pairs of reduced mass m . Each hydrogen bond is characterized by the stretching q^i and its conjugate momentum $p_i = m\dot{q}^i$. The elastic transverse force between neighboring pairs is tuned by the constant K , while the energy D and the inverse length a determine, respectively, the plateau and the narrowness of the on-site potential well that mimics the interaction between bases in each pair. It is understood that K , D , and a are all positive parameters. The transverse, external stress $h \geq 0$ is a computational tool useful in the evaluation of the susceptibility. Our interest in it lies in the fact that a phase transition can occur only when $h = 0$. We assume periodic boundary conditions.

The transfer operator technique [Dauxois *et al.* (2002)] maps the problem of computing the classical partition function into the easier task of evaluating the lowest energy eigenvalues of a ‘quantum’ mechanical Morse oscillator (no real quantum mechanics is involved, since the temperature plays the role of \hbar). One can then observe that, as the temperature increases, the number of levels belonging to the discrete spectrum decreases, until for some critical temperature $T_c = 2\sqrt{2KD}/(ak_B)$ only the continuous spectrum survives. This passage from a localized ground state to an

¹⁷The Peyrard–Bishop system has been proposed as a simple model for describing the DNA thermally induced denaturation [Grinza and Mossa (2004)].

unnormalizable one corresponds to the second-order phase transition of the statistical model. Various critical exponents can be analytically computed and all applicable scaling laws can be checked. The simplicity of this model permits an analytical computation of the largest Lyapunov exponent by exploiting the geometric method proposed in [Caiani *et al.* (1997)].

Mean-Field XY Hamiltonian System

The mean-field XY model describes a system of N equally coupled planar classical rotators (see [Antoni and Ruffo (1995); Casetti *et al.* (1999)]). It is defined by a Hamiltonian of the class (6.112) where the potential energy is

$$V(\varphi) = \frac{J}{2N} \sum_{i,j=1}^N [1 - \cos(\varphi_i - \varphi_j)] - h \sum_{i=1}^N \cos \varphi_i. \quad (6.120)$$

Here $\varphi_i \in [0, 2\pi]$ is the rotation angle of the i th rotator and h is an external field. Defining at each site i a classical spin vector $\mathbf{s}_i = (\cos \varphi_i, \sin \varphi_i)$ the model describes a planar (XY) Heisenberg system with interactions of equal strength among all the spins. We consider only the ferromagnetic case $J > 0$; for the sake of simplicity, we set $J = 1$. The equilibrium statistical mechanics of this system is exactly described, in the thermodynamic limit, by the *mean-field theory* [Antoni and Ruffo (1995)]. In the limit $h \rightarrow 0$, the system has a continuous phase transition, with classical critical exponents, at $T_c = 1/2$, or $\varepsilon_c = 3/4$, where $\varepsilon = E/N$ is the energy per particle.

The Lyapunov exponent λ_1 of this system is extremely sensitive to the phase transition. According to reported numerical simulations (see [Casetti *et al.* (1999)]), $\lambda_1(\varepsilon)$ is positive for $0 < \varepsilon < \varepsilon_c$, shows a sharp maximum immediately below the critical energy, and drops to zero at ε_c in the thermodynamic limit, where it remains zero in the whole region $\varepsilon > \varepsilon_c$, which corresponds to the thermodynamic disordered phase. In fact in this phase the system is integrable, reducing to an assembly of uncoupled rotators.

6.4.5.3 Euler Characteristics of Hamiltonian Systems

Recall that *Euler characteristic* χ is a number that is a characterization of the various classes of geometric figures based only on the topological relationship between the numbers of vertices V , edges E , and faces F , of a geometric Figure. This number, $\chi = F - E + V$, is the same for all figures the boundaries of which are composed of the same number of connected

pieces. Therefore, the Euler characteristic is a *topological invariant*, i.e., any two geometric figures that are homeomorphic to each other have the same Euler characteristic.

More specifically, a standard way to analyze a geometric Figure is to fragment it into other more familiar objects and then to examine how these pieces fit together. Take for example a surface M in the Euclidean 3D space. Slice M into pieces that are curved triangles (this is called a triangulation of the surface). Then count the number F of faces of the triangles, the number E of edges, and the number V of vertices on the tessellated surface. Now, no matter how we triangulate a compact surface Σ , its Euler characteristic, $\chi(\Sigma) = F - E + V$, will always equal a constant which is characteristic of the surface and which is invariant under diffeomorphisms $\phi : \Sigma \rightarrow \Sigma'$.

At higher dimensions this can be again defined by using higher dimensional generalizations of triangles (simplexes) and by defining the Euler characteristic $\chi(M)$ of the nD manifold M to be the alternating sum:

$$\begin{aligned} & \{\text{number of points}\} - \{\text{number of 2-simplices}\} + \\ & \{\text{number of 3-simplices}\} - \{\text{number of 4-simplices}\} + \dots \\ \text{i.e., } \chi(M) &= \sum_{k=0}^n (-1)^k (\text{number of faces of dimension } k). \end{aligned}$$

and then define the Euler characteristic of a manifold as the Euler characteristic of any simplicial complex homeomorphic to it. With this definition, circles and squares have Euler characteristic 0 and solid balls have Euler characteristic 1.

The Euler characteristic χ of a manifold is closely related to its *genus* g as $\chi = 2 - 2g$.¹⁸

Recall that a more standard topological definition of $\chi(M)$ is

$$\chi(M) = \sum_{k=0}^n (-1)^k b_k(M), \quad (6.121)$$

where b_k are the k th *Betti numbers* of M .

¹⁸Recall that the *genus* of a topological space such as a surface is a topologically invariant property defined as the largest number of nonintersecting simple closed curves that can be drawn on the surface without separating it, i.e., an integer representing the maximum number of cuts that can be made through it without rendering it disconnected. This is roughly equivalent to the number of holes in it, or handles on it. For instance: a point, line, and a sphere all have genus 0; a torus has genus 1, as does a coffee cup as a solid object (solid torus), a Möbius strip, and the symbol 0; the symbols 8 and B have genus 2; etc.

In general, it would be hopeless to try to practically calculate $\chi(M)$ from (6.121) in the case of non-trivial physical models at large dimension. Fortunately, there is a possibility given by the *Gauss–Bonnet formula*, that relates $\chi(M)$ with the total *Gauss–Kronecker curvature* of the manifold, (compare with (3.126) and (3.135))

$$\chi(M) = \gamma \int_M K_G d\sigma, \quad (6.122)$$

which is valid for even dimensional hypersurfaces of Euclidean spaces \mathbb{R}^N [here $\dim(M) = n \equiv N - 1$], and where:

$$\gamma = 2/\text{Vol}(S_1^n)$$

is twice the inverse of the volume of an n -dimensional sphere of unit radius S_1^n ; K_G is the Gauss–Kronecker curvature of the manifold;

$$d\sigma = \sqrt{\det(g)} dx^1 dx^2 \cdots dx^n$$

is the invariant volume measure of M and g is its Riemannian metric (induced from \mathbb{R}^N). Let us briefly sketch the meaning and definition of the Gauss–Kronecker curvature. The study of the way in which an n -surface M curves around in \mathbb{R}^N is measured by the way the normal direction changes as we move from point to point on the surface. The rate of change of the normal direction ξ at a point $x \in M$ in direction v is described by the *shape operator*

$$L_x(v) = -\mathcal{L}_v \xi = [v, \xi],$$

where v is a tangent vector at x and \mathcal{L}_v is the Lie derivative, hence

$$L_x(v) = -(\nabla \xi_1 \cdot v, \dots, \nabla \xi_{n+1} \cdot v);$$

gradients and vectors are represented in \mathbb{R}^N . As L_x is an operator of the tangent space at x into itself, there are n independent eigenvalues $\kappa_1(x), \dots, \kappa_n(x)$ which are called the principal curvatures of M at x [Thorpe (1979)]. Their product is the Gauss–Kronecker curvature:

$$K_G(x) = \prod_{i=1}^n \kappa_i(x) = \det(L_x).$$

Alternatively, recall that according to the *Morse theory*, it is possible to understand the topology of a given manifold by studying the regular critical points of a smooth *Morse function* defined on it. In our case, the

manifold M is the configuration space \mathbb{R}^N and the natural choice for the Morse function is the potential $V(q)$. Hence, one is lead to define the family M_v (6.114) of submanifolds of M .

A full characterization of the topological properties of M_v generally requires the critical points of $V(q)$, which means solving the equations

$$\partial_{q^i} V = 0, \quad (i = 1, \dots, N). \quad (6.123)$$

Moreover, one has to calculate the indexes of all the critical points, that is the number of negative eigenvalues of the Hessian $\partial^2 V / (\partial q^i \partial q_j)$. Then the Euler characteristic $\chi(M_v)$ can be computed by means of the formula

$$\chi(M_v) = \sum_{k=0}^N (-1)^k \mu_k(M_v), \quad (6.124)$$

where $\mu_k(M_v)$ is the total number of critical points of $V(q)$ on M_v which have index k , i.e., the so-called *Morse numbers* of a manifold M , which happen to be upper bounds of the Betti numbers,

$$b_k(M) \leq \mu_k(M) \quad (k = 0, \dots, n). \quad (6.125)$$

Among all the Morse functions on a manifold M , there is a special class, called *perfect Morse functions*, for which the Morse inequalities (6.125) hold as equalities. Perfect Morse functions characterize completely the topology of a manifold.

Now, we continue with our two examples started before.

Peyrard–Bishop System. If applied to any generic model, calculation of (6.124) turns out to be quite formidable, but the exceptional simplicity of the Peyrard–Bishop model (6.119) makes it possible to carry on completely the topological analysis without invoking equation (6.124).

For the potential in exam, equation (6.123) results in the nonlinear system

$$\frac{a}{R}(q^{i+1} - 2q^i + q^{i-1}) = h - 2(e^{-2aq^i} - e^{-aq^i}),$$

where $R = Da^2/K$ is a dimensionless ratio. It is easy to verify that a particular solution is given by

$$q^i = -\frac{1}{a} \ln \frac{1 + \sqrt{1 + 2h}}{2}, \quad (i = 1, \dots, N).$$

The corresponding minimum of potential energy is

$$V_{\min} = ND \left(\frac{1+h-\sqrt{1+2h}}{2} - h \ln \frac{1+\sqrt{1+2h}}{2} \right).$$

Mean-Field XY Model. In the case of the mean-field XY model (6.120) it is possible to show analytically that a topological change in the configuration space exists and that it can be related to the thermodynamic phase transition. Consider again the family M_v of submanifolds of the configuration space defined in (6.114); now the potential energy per degree of freedom is that of the mean-field XY model, i.e.,

$$\mathcal{V}(\varphi) = \frac{V(\varphi)}{N} = \frac{J}{2N^2} \sum_{i,j=1}^N [1 - \cos(\varphi_i - \varphi_j)] - h \sum_{i=1}^N \cos \varphi_i,$$

where $\varphi_i \in [0, 2\pi]$. Such a function can be considered a Morse function on M , so that, according to Morse theory, all these manifolds have the same topology until a critical level $\mathcal{V}^{-1}(v_c)$ is crossed, where the topology of M_v changes.

A change in the topology of M_v can only occur when v passes through a critical value of \mathcal{V} . Thus in order to detect topological changes in M_v we have to find the critical values of \mathcal{V} , which means solving the equations

$$\partial \varphi_i \mathcal{V}(\varphi) = 0, \quad (i = 1, \dots, N). \quad (6.126)$$

For a general potential energy function \mathcal{V} , the solution of (6.126) would be a formidable task, but in the case of the mean-field XY model, the mean-field character of the interaction greatly simplifies the analysis, allowing an analytical treatment of (6.126); moreover, a projection of the configuration space onto a 2D plane is possible [Casetti *et al.* (1999); Casetti *et al.* (2003)].

6.4.6 Application: *Force-Field Psychodynamics*

In this section, which is written in the fashion of the above *quantum brain modelling* (see subsection 6.3.8 above), we present the top level of natural biodynamics, using geometrical generalization of the *Feynman path integral*. To formulate the basics of *force-field psychodynamics*, we use the *action-amplitude picture* of the $BODY \rightleftharpoons MIND$ adjunction:

↓ **Deterministic (causal) world of *Human BODY*** ↓

$$Action : S[q^n] = \int_{t_{in}}^{t_{out}} (E_k - E_p + Wrk + Src^\pm) dt$$

$$Amplitude : \langle out|in \rangle = \oint \mathcal{D}[w_n q^n] e^{iS[q^n]}$$

↑ **Probabilistic (fuzzy) world of *Human MIND*** ↑

In the action integral, E_k, E_p, Wrk and Src^\pm denote the kinetic end potential energies, work done by dissipative/driving forces and other energy sources/sinks, respectively. In the amplitude integral, the peculiar sign \oint denotes integration along smooth paths and summation along discrete Markov chains; i is the imaginary unit, w_n are synaptic-like weights, while \mathcal{D} is the Feynman path differential (defined below) calculated along the configuration trajectories q^n . The action $S[q^n]$, through the *least action principle* $\delta S = 0$, leads to all biodynamic equations considered so far (in generalized Lagrangian and Hamiltonian form). At the same time, the action $S[q^n]$ figures in the exponent of the path integral \oint , defining the probability transition amplitude $\langle out|in \rangle$. In this way, the whole body dynamics is incorporated in the mind dynamics. This *adaptive path integral* represents an *infinite-dimensional neural network*, suggesting an infinite capacity of human brain/mind.

For a long time the cortical systems for *language and actions* were believed to be independent modules. However, according to the recent research of [Pulvermüller (2005)], as these systems are reciprocally connected with each other, information about language and actions might interact in distributed neuronal assemblies. A critical case is that of action words that are semantically related to different parts of the body (e.g. ‘pick’, ‘kick’, ‘lick’,...). The author suggests that the comprehension of these words might specifically, rapidly and automatically activate the motor system in a somatotopic manner, and that their comprehension rely on activity in the action system.

6.4.6.1 *Motivational Cognition in the Life Space Foam*

Applications of nonlinear dynamical systems (NDS) theory in psychology have been encouraging, if not universally productive/effective [Met-

zger (1997)]. Its historical antecedents can be traced back to Piaget's [Piaget *et. al.* (1992)] and Vygotsky's [Vygotsky (1982)] interpretations of the dynamic relations between action and thought, Lewinian theory of social dynamics and cognitive-affective development [Lewin (1951); Gold (1999)], and Bernstein's [Bernstein (1947)] theory of self-adjusting, goal-driven motor action.

Now, both the original *Lewinian force-field theory* in psychology (see [Lewin (1951); Gold (1999)]) and modern decision-field dynamics (see [Busemeyer and Townsend (1993); Roe *et al.* (2001); Busemeyer and Diederich (2002)]) are based on the classical Lewinian concept of an individual's *life space*.¹⁹ As a topological construct, Lewinian life space represents a person's psychological environment that contains *regions* separated by dynamical permeable *boundaries*. As a field construct, on the other hand, the life space is not empty: each of its regions is characterized by *valence* (ranging from positive or negative and resulting from an interaction between the person's *needs* and the dynamics of their *environment*). Need is an energy construct, according to Lewin. It creates *tension* in the person, which, in combination with other tensions, initiates and sustains behavior. Needs vary from the most primitive urges to the most idiosyncratic intentions and can be both internally generated (e.g., thirst or hunger) and stimulus-induced (e.g., an urge to buy something in response to a TV advertisement). Valences are, in essence, personal values dynamically derived from the person's needs and attached to various regions in their life space. As a field, the life space generates forces pulling the person towards positively-valenced regions and pushing them away from regions with negative valence. Lewin's term for these forces is *vectors*. Combinations of multiple vectors in the life space cause the person to move from one region towards another. This movement is termed *locomotion* and it may range from overt behavior to cognitive shifts (e.g., between alternatives in a decision-making process). Locomotion normally results in crossing the boundaries between regions. When their permeability is degraded, these boundaries become *barriers* that restrain locomotion. Life space model, thus, offers a meta-theoretical language to describe a wide range of behaviors, from goal-directed action to intrapersonal conflicts and multi-alternative decision-making.

In order to formalize the Lewinian life-space concept, a set of *action principles* need to be associated to Lewinian force-fields, (loco)motion

¹⁹The work presented in this subsection has been developed in collaboration with Dr. Eugene Aidman, Senior Research Scientist, Human Systems Integration, Land Operations Division, Defence Science & Technology Organisation, Australia.

paths (representing mental abstractions of biomechanical paths [Ivancevic (2004)]) and life space geometry. As an extension of the Lewinian concept, in this paper we introduce a new concept of *life-space foam* (LSF, see Figure 6.12). According to this new concept, Lewin's life space can be represented as a *geometrical functor* with globally smooth macro-dynamics, which is at the same time underpinned by wildly fluctuating, non-smooth, local micro-dynamics, describable by *Feynman's*: (i) *sum-over-histories* \int_{paths} , (ii) *sum-over-fields* \int_{fields} , and (iii) *sum-over-geometries* \int_{geom} .

LSF is thus a two-level *geometro-dynamical functor*, representing these two distinct types of dynamics within the Lewinian life space. At its *macroscopic spatio-temporal level*, LSF appears as a 'nice & smooth' geometrical functor with globally predictable dynamics – formally, a smooth n -dimensional manifold M with local Riemannian metrics $g_{ij}(x)$, smooth force-fields and smooth (loco)motion paths, as conceptualized in the Lewinian theory. To model the global and smooth macro-level LSF-paths, fields and geometry, we use the general physics-like *principle of the least action*.

Now, the apparent smoothness of the macro-level LSF is achieved by the existence of another level underneath it. This *micro-level* LSF is actually a collection of wildly fluctuating force-fields, (loco)motion paths, curved regional geometries and topologies with holes. The micro-level LSF is proposed as an extension of the Lewinian concept: it is characterized by uncertainties and fluctuations, enabled by microscopic time-level, microscopic transition paths, microscopic force-fields, local geometries and varying topologies with holes. To model these fluctuating microscopic LSF-structures, we use three instances of *adaptive path integral*, defining a multi-phase and multi-path (also multi-field and multi-geometry) *transition* process from *intention* to the goal-driven *action*.

We use the new LSF concept to develop modelling framework for motivational dynamics (MD) and induced cognitive dynamics (CD).

According to Heckhausen (see [Heckhausen (1977)]), *motivation* can be thought of as a process of *energizing* and *directing the action*. The process of energizing can be represented by Lewin's *force-field analysis* and Vygotsky's *motive formation* (see [Vygotsky (1982); Aidman and Leontiev (1991)]), while the process of directing can be represented by *hierarchical action control* (see [Bernstein (1947); Bernstein (1935); Kuhl (1985)]).

Motivation processes both precede and coincide with every goal-directed

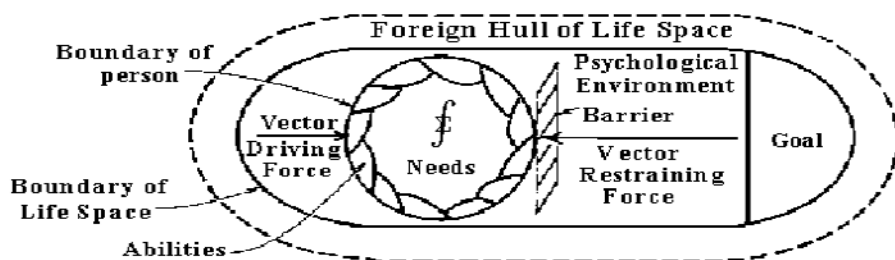


Fig. 6.12 Diagram of the *life space foam*: Lewinian life space with an adaptive path integral acting inside it and generating microscopic fluctuation dynamics.

action. Usually these motivation processes include the sequence of the following four feedforward *phases* [Vygotsky (1982); Aidman and Leontiev (1991)]: (*)

- (1) *Intention Formation* \mathcal{F} , including: decision making, commitment building, etc.
- (2) *Action Initiation* \mathcal{I} , including: handling conflict of motives, resistance to alternatives, etc.
- (3) *Maintaining the Action* \mathcal{M} , including: resistance to fatigue, distractions, etc.
- (4) *Termination* \mathcal{T} , including parking and avoiding addiction, i.e., staying in control.

With each of the phases $\{\mathcal{F}, \mathcal{I}, \mathcal{M}, \mathcal{T}\}$ in (*), we can associate a *transition propagator* – an ensemble of (possibly crossing) feedforward paths propagating through the ‘wood of obstacles’ (including topological holes in the LSF, see Figure 6.13), so that the complete *transition functor* $\mathcal{T}A$ is a product of propagators (as well as sum over paths). All the phases–propagators are controlled by a unique *Monitor* feedback process.

In this subsection we propose an *adaptive path integral* formulation for the motivational–transition functor $\mathcal{T}A$. In essence, we sum/integrate over different paths and make a product (composition) of different phases–propagators. Recall that this is the most general description of the general *Markov stochastic process*.

We will also attempt to demonstrate the utility of the same LSF–formalisms in representing cognitive functions, such as memory, learning and decision making. For example, in the classical *Stimulus encoding* \rightarrow *Search* \rightarrow *Decision* \rightarrow *Response* sequence [Sternberg (1969);

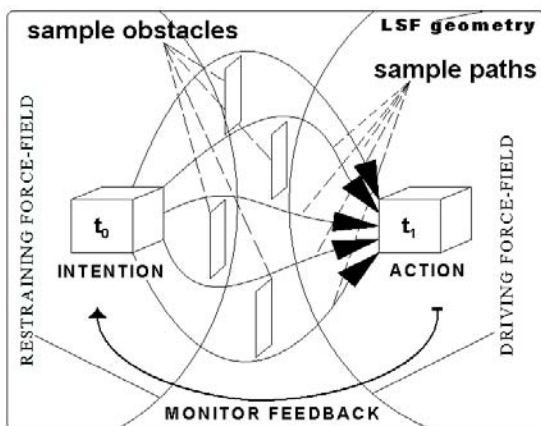


Fig. 6.13 *Transition-propagator* corresponding to each of the motivational phases $\{\mathcal{F}, \mathcal{I}, \mathcal{M}, \mathcal{T}\}$, consisting of an ensemble of feedforward paths propagating through the ‘wood of obstacles’. The paths affected by driving and restraining force-fields, as well as by the local LSF-geometry. Transition goes from *Intention*, occurring at a sample time instant t_0 , to *Action*, occurring at some later time t_1 . Each propagator is controlled by its own *Monitor* feedback. All together they form the transition functor \mathcal{TA} .

Ashcraft (1994)], the environmental input-triggered *sensory memory* and *working memory* (WM) can be interpreted as operating at the micro-level force-field under the executive control of the *Monitor* feedback, whereas *search* can be formalized as a *control* mechanism guiding retrieval from the long-term memory (LTM, itself shaped by learning) and filtering material relevant to decision making into the WM. The essential measure of these mental processes, the *processing speed* (essentially determined by Sternberg’s reaction-time) can be represented by our (loco)motion speed \dot{x} .

Six Faces of the Life Space Foam

The LSF has three forms of appearance: *paths + field + geometries*, acting on both macro-level and micro-level, which is six modes in total. In this section, we develop three least action principles for the macro-LSF-level and three adaptive path integrals for the micro-LSF-level. While developing our psycho-physical formalism, we will address the behavioral issues of motivational fatigue, learning, memory and decision making.

General Formalism

At both macro- and micro-levels, the total LSF represents a union of

transition paths, force-fields and geometries, formally written as

$$\begin{aligned} LSF_{total} &:= LSF_{paths} \bigcup LSF_{fields} \bigcup LSF_{geom} \\ &\equiv \oint_{paths} + \oint_{fields} + \oint_{geom}. \end{aligned} \quad (6.127)$$

Corresponding to each of the three LSF-subspaces in (6.127) we formulate:

- (1) The *least action principle*, to model deterministic and predictive, macro-level MD & CD, giving a unique, global, causal and smooth path-field-geometry on the macroscopic spatio-temporal level; and
- (2) Associated *adaptive path integral* to model uncertain, fluctuating and probabilistic, micro-level MD & CD, as an ensemble of local paths-fields-geometries on the microscopic spatio-temporal level, to which the global macro-level MD & CD represents both time and ensemble *average* (which are equal according to the *ergodic hypothesis*).

In the proposed formalism, transition paths $x^i(t)$ are affected by the force-fields $\varphi^k(t)$, which are themselves affected by geometry with metric g_{ij} .

Global Macro-Level of LSF_{total} . In general, at the *macroscopic* LSF-level we first formulate the *total action* $S[\Phi]$, the central quantity in our formalism that has psycho-physical dimensions of $Energy \times Time = Effort$, with immediate cognitive and motivational applications: *the greater the action – the higher the speed of cognitive processes and the lower the macroscopic fatigue* (which includes all sources of physical, cognitive and emotional fatigue that influence motivational dynamics). The action $S[\Phi]$ depends on macroscopic paths, fields and geometries, commonly denoted by an abstract field symbol Φ^i . The action $S[\Phi]$ is formally defined as a temporal integral from the *initial* time instant t_{ini} to the *final* time instant t_{fin} ,

$$S[\Phi] = \int_{t_{ini}}^{t_{fin}} \mathfrak{L}[\Phi] dt, \quad (6.128)$$

with *Lagrangian density* given by

$$\mathfrak{L}[\Phi] = \int d^n x \mathcal{L}(\Phi_i, \partial_{x^j} \Phi^i),$$

where the integral is taken over all n coordinates $x^j = x^j(t)$ of the LSF, and $\partial_{x^j} \Phi^i$ are time and space partial derivatives of the Φ^i -variables over coordinates.

Second, we formulate the *least action principle* as a minimal variation δ of the action $S[\Phi]$

$$\delta S[\Phi] = 0, \quad (6.129)$$

which, using techniques from the calculus of variations gives, in the form of the so-called Euler–Lagrangian equations, a shortest (loco)motion path, an extreme force–field, and a life–space geometry of minimal curvature (and without holes). In this way, we effectively derive a *unique globally smooth transition functor*

$$\mathcal{TA} : INTENTION_{t_{ini}} \Rightarrow ACTION_{t_{fin}}, \quad (6.130)$$

performed at a macroscopic (global) time–level from some initial time t_{ini} to the final time t_{fin} .

In this way, we get macro–objects in the global LSF: a single path described Newtonian–like equation of motion, a single force–field described by Maxwellian–like field equations, and a single obstacle–free Riemannian geometry (with global topology without holes).

For example, recall that in the period 1945–1949 J. Wheeler and R. Feynman developed their *action-at-a-distance electrodynamics* [Wheeler and Feynman (1949)], in complete experimental agreement with the classical Maxwell’s electromagnetic theory, but at the same time avoiding the complications of divergent self–interaction of the Maxwell’s theory as well as eliminating its infinite number of field degrees of freedom. In Wheeler–Feynman view, “Matter consists of electrically charged particles,” so they found a form for the action directly involving the motions of the charges only, which upon variation would give the Newtonian–like equations of motion of these charges. Here is the expression for this action in the flat space–time, which is in the core of quantum electrodynamics:

$$S[x; t_i, t_j] = \frac{1}{2} m_i \int (\dot{x}_\mu^i)^2 dt_i + \frac{1}{2} e_i e_j \int \int \delta(I_{ij}^2) \dot{x}_\mu^i(t_i) \dot{x}_\mu^j(t_j) dt_i dt_j$$

with

$$I_{ij}^2 = [x_\mu^i(t_i) - x_\mu^j(t_j)] [x_\mu^i(t_i) - x_\mu^j(t_j)], \quad (6.131)$$

where $x_\mu^i = x_\mu^i(t_i)$ is the four–vector position of the i th particle as a function of the proper time t_i , while $\dot{x}_\mu^i(t_i) = dx_\mu^i/dt_i$ is the velocity four–vector. The first term in the action (6.131) is the ordinary mechanical action in Euclidean space, while the second term defines the electrical interaction of

the charges, representing the Maxwell-like field (it is summed over each pair of charges; the factor $\frac{1}{2}$ is to count each pair once, while the term $i = j$ is omitted to avoid self-action; the interaction is a double integral over a delta function of the square of space-time interval I^2 between two points on the paths; thus, interaction occurs only when this interval vanishes, that is, along light cones [Wheeler and Feynman (1949)]).

Now, from the point of view of Lewinian geometrical force-fields and (loco)motion paths, we can give the following life-space interpretation to the Wheeler-Feynman action (6.131). The mechanical-like locomotion term occurring at the single time t , needs a covariant generalization from the flat 4D Euclidean space to the n D smooth Riemannian manifold, so it becomes (see e.g., [Ivancevic (2004)])

$$S[x] = \frac{1}{2} \int_{t_{ini}}^{t_{fin}} g_{ij} \dot{x}^i \dot{x}^j dt,$$

where g_{ij} is the Riemannian metric tensor that generates the total ‘kinetic energy’ of (loco)motions in the life space.

The second term in (6.131) gives the sophisticated definition of Lewinian force-fields that drive the psychological (loco)motions, if we interpret electrical charges e_i occurring at different times t_i as motivational charges – needs.

Local Micro-Level of LSF_{total} . After having properly defined macro-level MD & CD, with a unique transition map F (including a unique motion path, driving field and smooth geometry), we move down to the *microscopic* LSF-level of rapidly fluctuating MD & CD, where we cannot define a unique and smooth path-field-geometry. The most we can do at this level of *fluctuating uncertainty*, is to formulate an adaptive path integral and calculate overall probability amplitudes for ensembles of local transitions from one LSF-point to the neighboring one. This *probabilistic transition micro-dynamics* functor is defined by a multi-path (field and geometry, respectively) and multi-phase *transition amplitude* $\langle \text{Action} | \text{Intention} \rangle$ of corresponding to the globally-smooth transition map (6.130). This absolute square of this probability amplitude gives the *transition probability* of occurring the final state of *Action* given the initial state of *Intention*,

$$P(\text{Action} | \text{Intention}) = |\langle \text{Action} | \text{Intention} \rangle|^2.$$

The total transition amplitude from the state of *Intention* to the state of

Action is defined on LSF_{total}

$$\mathcal{TA} \equiv \langle Action | Intention \rangle_{total} : INTENTION_{t_0} \Rightarrow ACTION_{t_1}, \quad (6.132)$$

given by adaptive generalization of the Feynman's path integral [Feynman and Hibbs (1965); Feynman (1972); Feynman (1998)]. The transition map (6.132) calculates the *overall probability amplitude* along a multitude of wildly fluctuating paths, fields and geometries, performing the *microscopic* transition from the micro-state $INTENTION_{t_0}$ occurring at initial micro-time instant t_0 to the micro-state $ACTION_{t_1}$ at some later micro-time instant t_1 , such that all micro-time instants fit inside the global transition interval $t_0, t_1, \dots, t_s \in [t_{ini}, t_{fin}]$. It is symbolically written as

$$\langle Action | Intention \rangle_{total} := \oint \mathcal{D}[w\Phi] e^{iS[\Phi]}, \quad (6.133)$$

where the Lebesgue integration is performed over all continuous $\Phi_{con}^i = paths + field + geometries$, while summation is performed over all discrete processes and regional topologies Φ_{dis}^j). The symbolic differential $\mathcal{D}[w\Phi]$ in the general path integral (6.133), represents an *adaptive path measure*, defined as a weighted product

$$\mathcal{D}[w\Phi] = \lim_{N \rightarrow \infty} \prod_{s=1}^N w_s d\Phi_s^i, \quad (i = 1, \dots, n = con + dis), \quad (6.134)$$

which is in practice satisfied with a large N corresponding to infinitesimal temporal division of the four motivational phases (*). Technically, the path integral (6.133) calculates the *amplitude* for the transition functor $\mathcal{TA} : Intention \Rightarrow Action$.

In the exponent of the path integral (6.133) we have the action $S[\Phi]$ and the imaginary unit $i = \sqrt{-1}$ (i can be converted into the real number -1 using the so-called *Wick rotation*, see next subsection).

In this way, we get a range of micro-objects in the local LSF at the short time-level: ensembles of rapidly fluctuating, noisy and crossing paths, force-fields, local geometries with obstacles and topologies with holes. However, by averaging process, both in time and along ensembles of paths, fields and geometries, we recover the corresponding global MD & CD variables.

Infinite-Dimensional Neural Network. The adaptive path integral (6.133) incorporates the *local learning process* according to the standard formula: *New Value = Old Value + Innovation*. The general *weights* $w_s = w_s(t)$ in (6.134) are updated by the *MONITOR* feedback during the

transition process, according to one of the two standard neural learning schemes, in which the micro-time level is traversed in discrete steps, i.e., if $t = t_0, t_1, \dots, t_s$ then $t + 1 = t_1, t_2, \dots, t_{s+1}$:

- (1) A *self-organized, unsupervised* (e.g., Hebbian-like [Hebb (1949)]) learning rule:

$$w_s(t+1) = w_s(t) + \frac{\sigma}{\eta}(w_s^d(t) - w_s^a(t)), \quad (6.135)$$

where $\sigma = \sigma(t)$, $\eta = \eta(t)$ denote *signal* and *noise*, respectively, while superscripts d and a denote *desired* and *achieved* micro-states, respectively; or

- (2) A certain form of a *supervised gradient descent learning*:

$$w_s(t+1) = w_s(t) - \eta \nabla J(t), \quad (6.136)$$

where η is a small constant, called the *step size*, or the *learning rate* and $\nabla J(n)$ denotes the gradient of the ‘performance hyper-surface’ at the t -th iteration.

Both Hebbian and supervised learning are used for the local decision making process (see below) occurring at the intention formation phase \mathcal{F} .

In this way, local micro-level of LSF_{total} represents an infinite-dimensional neural network. In the cognitive psychology framework, our adaptive path integral (6.133) can be interpreted as *semantic integration* (see [Bransford and Franks (1971); Ashcraft (1994)]).

Motion and Decision Making in LSF_{paths}

On the macro-level in the subspace LSF_{paths} we have the (loco)*motion action principle*

$$\delta S[x] = 0,$$

with the *Newtonian-like action* $S[x]$ given by

$$S[x] = \int_{t_{ini}}^{t_{fin}} dt \left[\frac{1}{2} g_{ij} \dot{x}^i \dot{x}^j + \varphi^i(x^i) \right], \quad (6.137)$$

where overdot denotes time derivative, so that \dot{x}^i represents *processing speed*, or (loco)motion velocity vector. The first bracket term in (6.137) represents the kinetic energy T ,

$$T = \frac{1}{2} g_{ij} \dot{x}^i \dot{x}^j,$$

generated by the *Riemannian metric tensor* g_{ij} , while the second bracket term, $\varphi^i(x^i)$, denotes the family of potential force-fields, driving the (loco)mo-tions $x^i = x^i(t)$ (the *strengths* of the fields $\varphi^i(x^i)$ depend on their positions x^i in LSF, see LSF_{fields} below). The corresponding Euler–Lagrangian equation gives the Newtonian-like equation of motion

$$\frac{d}{dt}T_{\dot{x}^i} - T_{x^i} = -\varphi^i_{x^i}, \quad (6.138)$$

(subscripts denote the partial derivatives), which can be put into the standard Lagrangian form

$$\frac{d}{dt}L_{\dot{x}^i} = L_{x^i}, \quad \text{with} \quad L = T - \varphi^i(x^i).$$

In the next subsection we use the micro-level implications of the action $S[x]$ as given by (6.137), for dynamical descriptions of the local decision-making process.

On the micro-level in the subspace LSF_{paths} , instead of a single path defined by the Newtonian-like equation of motion (6.138), we have an ensemble of fluctuating and crossing paths with weighted probabilities (of the unit total sum). This ensemble of micro-paths is defined by the simplest instance of our adaptive path integral (6.133), similar to the Feynman’s original *sum over histories*,

$$\langle \text{Action} | \text{Intention} \rangle_{paths} = \oint \mathcal{D}[wx] e^{iS[x]}, \quad (6.139)$$

where $\mathcal{D}[wx]$ is a functional measure on the *space of all weighted paths*, and the exponential depends on the action $S[x]$ given by (6.137). This procedure can be redefined in a mathematically cleaner way if we Wick-rotate the time variable t to imaginary values $t \mapsto \tau = it$, thereby making all integrals real:

$$\oint \mathcal{D}[wx] e^{iS[x]} \xrightarrow{\text{Wick}} \oint \mathcal{D}[wx] e^{-S[x]}. \quad (6.140)$$

Discretization of (6.140) gives the *thermodynamic-like partition function*

$$Z = \sum_j e^{-w_j E^j / T}, \quad (6.141)$$

where E^j is the motion energy eigenvalue (reflecting each possible motivational energetic state), T is the temperature-like environmental control parameter, and the sum runs over all motion energy eigenstates (labelled

by the index j). From (6.141), we can further calculate all thermodynamic-like and statistical properties of MD & CD (see e.g., [Feynman (1972)]), as for example, *transition entropy* $S = k_B \ln Z$, etc.

From cognitive perspective, our adaptive path integral (6.139) calculates all (alternative) pathways of information flow during the transition $Intention \rightarrow Action$.

In the language of transition-propagators, the integral over histories (6.139) can be decomposed into the product of propagators (i.e., Fredholm kernels or Green functions) corresponding to the cascade of the four motivational phases (*)

$$\langle Action | Intention \rangle_{paths} = \int dx^{\mathcal{F}} dx^{\mathcal{I}} dx^{\mathcal{M}} dx^{\mathcal{T}} K(\mathcal{F}, \mathcal{I}) K(\mathcal{I}, \mathcal{M}) K(\mathcal{M}, \mathcal{T}), \quad (6.142)$$

satisfying the Schrödinger-like equation (see e.g., [Dirac (1982)])

$$i \partial_t \langle Action | Intention \rangle_{paths} = H_{Action} \langle Action | Intention \rangle_{paths}, \quad (6.143)$$

where H_{Action} represents the Hamiltonian (total energy) function available at the state of *Action*. Here our ‘golden rule’ is: the higher the H_{Action} , the lower the microscopic fatigue.

In the connectionist language, our propagator expressions (6.142–6.143) represent *activation dynamics*, to which our *Monitor* process gives a kind of *backpropagation* feedback, a version of the basic supervised learning (6.136).

Mechanisms of Decision-Making under Uncertainty. The basic question about our local decision making process, occurring under uncertainty at the intention formation phase \mathcal{F} , is: Which alternative to choose? (see [Roe *et al.* (2001); Grossberg (1982); Grossberg (1999); Grossberg (1988); Ashcraft (1994)]). In our path-integral language this reads: Which path (alternative) should be given the highest probability weight w ? Naturally, this problem is iteratively solved by the learning process (6.135–6.136), controlled by the *MONITOR* feedback, which we term *algorithmic approach*.

In addition, here we analyze qualitative mechanics of the local decision making process under uncertainty, as a *heuristic approach*. This qualitative analysis is based on the micro-level interpretation of the Newtonian-like action $S[x]$, given by (6.137) and figuring both processing speed \dot{x} and LTM (i.e., the force-field $\varphi(x)$, see next subsection). Here we consider three different cases:

- (1) If the potential $\varphi(x)$ is not very dependent upon position $x(t)$, then the more direct paths contribute the most, as longer paths, with higher mean square velocities $[\dot{x}(t)]^2$ make the exponent more negative (after Wick rotation (6.140)).
- (2) On the other hand, suppose that $\varphi(x)$ does indeed depend on position x . For simplicity, let the potential increase for the larger values of x . Then a direct path does not necessarily give the largest contribution to the overall transition probability, because the integrated value of the potential is higher than over another paths.
- (3) Finally, consider a path that deviates widely from the direct path. Then $\varphi(x)$ decreases over that path, but at the same time the velocity \dot{x} increases. In this case, we expect that the increased velocity \dot{x} would more than compensate for the decreased potential over the path.

Therefore, the most important path (i.e., the path with the highest weight w) would be one for which any smaller integrated value of the surrounding field potential $\varphi(x)$ is more than compensated for by an increase in kinetic-like energy $\frac{m}{2}\dot{x}^2$. In principle, this is neither the most direct path, nor the longest path, but rather a middle way between the two. Formally, it is the path along which the average Lagrangian is minimal,

$$< \frac{m}{2}\dot{x}^2 + \varphi(x) > \longrightarrow \min, \quad (6.144)$$

i.e., the *path that requires minimal memory* (both LTM and WM, see LSF_{fields} below) and *processing speed*. This mechanical result is consistent with the ‘filter theory’ of *selective attention* [Broadbent (1958)], proposed in an attempt to explain a range of the existing experimental results. This theory postulates a low level filter that allows only a limited number of percepts to reach the brain at any time. In this theory, the importance of conscious, directed attention is minimized. The type of attention involving low level filtering corresponds to the concept of *early selection* [Broadbent (1958)].

Although we termed this ‘heuristic approach’ in the sense that we can instantly feel both the processing speed \dot{x} and the LTM field $\varphi(x)$ involved, there is clearly a psycho-physical rule in the background, namely the averaging minimum relation (6.144).

From the decision making point of view, all possible paths (alternatives) represent the *consequences* of decision making. They are, by default, *short-term consequences*, as they are modelled in the micro-time-level. However, the path integral formalism allows calculation of the *long-term*

consequences, just by extending the integration time, $t_{fin} \rightarrow \infty$. Besides, this *averaging decision mechanics* – choosing the optimal path – actually performs the ‘averaging lift’ in the LSF: from micro– to the macro–level.

Force–Fields and Memory in LSF_{fields}

At the macro–level in the subspace LSF_{fields} we formulate the *force–field action principle*

$$\delta S[\varphi] = 0, \quad (6.145)$$

with the action $S[\varphi]$ dependent on Lewinian force–fields $\varphi^i = \varphi^i(x)$ ($i = 1, \dots, N$), defined as a temporal integral

$$S[\varphi] = \int_{t_{ini}}^{t_{fin}} \mathfrak{L}[\varphi] dt, \quad (6.146)$$

with Lagrangian density given by

$$\mathfrak{L}[\varphi] = \int d^n x \mathcal{L}(\varphi_i, \partial_{x^j} \varphi^i),$$

where the integral is taken over all n coordinates $x^j = x^j(t)$ of the LSF, and $\partial_{x^j} \varphi^i$ are partial derivatives of the field variables over coordinates.

On the micro–level in the subspace LSF_{fields} we have the Feynman–type *sum over fields* φ^i ($i = 1, \dots, N$) given by the adaptive path integral

$$\langle \text{Action} | \text{Intention} \rangle_{fields} = \oint \mathcal{D}[w\varphi] e^{iS[\varphi]} \xrightarrow{\text{Wick}} \oint \mathcal{D}[w\varphi] e^{-S[\varphi]}, \quad (6.147)$$

with action $S[\varphi]$ given by temporal integral (6.146). (Choosing special forms of the force–field action $S[\varphi]$ in (6.147) defines micro–level MD & CD, in the LSF_{fields} space, that is similar to standard quantum–field equations, see e.g., [Ramond (1990)].) The corresponding partition function has the form similar to (6.141), but with field energy levels.

Regarding topology of the force fields, we have in place *n–categorical Lagrangian–field structure* on the Riemannian LSF manifold M ,

$$\Phi^i : [0, 1] \rightarrow M, \quad \Phi^i : \Phi_0^i \mapsto \Phi_1^i,$$

generalized from the recursive homotopy dynamics (3.205) above, using

$$\begin{aligned} \frac{d}{dt} f_{\dot{x}^i} &= f_{x^i} \longrightarrow \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial_\mu \Phi^i} \right) = \frac{\partial \mathcal{L}}{\partial \Phi^i}, \\ \text{with } [x_0, x_1] &\longrightarrow [\Phi_0^i, \Phi_1^i]. \end{aligned}$$

Relationship between Memory and Force-Fields. As already mentioned, the subspace LSF_{fields} is related to our *memory storage* [Ashcraft (1994)]. Its global macro-level represents the *long-term memory* (LTM), defined by the least action principle (6.145), related to *cognitive economy* in the model of *semantic memory* [Ratcliff (1978); Collins and Quillian (1969)]. Its local micro-level represents *working memory* (WM), a limited-capacity ‘bottleneck’ defined by the adaptive path integral (6.147). According to our formalism, each of Miller’s 7 ± 2 units [Miller (1956)] of the local WM are adaptively stored and averaged to give the global LTM capacity (similar to the physical notion of potential). This averaging memory lift, from WM to LTM represents *retroactive interference*, while the opposite direction, given by the path integral (6.147) itself, represents *proactive interference*. Both retroactive and proactive interferences are examples of the impact of cognitive contexts on memory. Motivational contexts can exert their influence, too. For example, a reduction in task-related recall following the completion of the task is one of the clearest examples of force-field influences on memory: the amount of details remembered of a task declines as the force-field tension to complete the task is reduced by actually completing it.

Once defined, the global LTM potential $\varphi = \varphi(x)$ is then affecting the locomotion transition paths through the path action principle (6.137), as well as general learning (6.135–6.136) and decision making process (6.144).

On the other hand, the two levels of LSF_{fields} fit nicely into the two levels of processing framework, as presented by [Craik and Lockhart (1972)], as an alternative to theories of separate stages for sensory, working and long-term memory. According to the *levels of processing framework*, stimulus information is processed at multiple levels simultaneously depending upon its characteristics. In this framework, our macro-level memory field, defined by the fields action principle (6.145), corresponds to the *shallow memory*, while our micro-level memory field, defined by the adaptive path integral (6.147), corresponds to the *deep memory*.

Geometries, Topologies and Noise in LSF_{geom}

On the macro-level in the subspace LSF_{geom} representing an n -dimensional smooth manifold M with the global Riemannian metric tensor g_{ij} , we formulate the *geometrical action principle*

$$\delta S[g_{ij}] = 0,$$

where $S = S[g_{ij}]$ is the n -dimensional *geodesic action* on M ,

$$S[g_{ij}] = \int d^n x \sqrt{g_{ij} dx^i dx^j}. \quad (6.148)$$

The corresponding Euler–Lagrangian equation gives the *geodesic equation* of the *shortest path* in the manifold M ,

$$\ddot{x}^i + \Gamma_{jk}^i \dot{x}^j \dot{x}^k = 0,$$

where the symbol Γ_{jk}^i denotes the so-called *affine connection* which is the source of *curvature*, which is geometrical description for *noise* (see [Ingber (1997); Ingber (1998)]). The higher the local curvatures of the LSF-manifold M , the greater the noise in the life space. This noise is the source of our micro-level fluctuations. It can be internal or external; in both cases it curves our micro-LSF.

Otherwise, if instead we choose an n -dimensional Hilbert-like action (see [Misner *et al.* (1973)]),

$$S[g_{ij}] = \int d^n x \sqrt{\det |g_{ij}|} R, \quad (6.149)$$

where R is the scalar curvature (derived from Γ_{jk}^i), we get the n -dimensional Einstein-like equation: $G_{ij} = 8\pi T_{ij}$, where G_{ij} is the Einstein-like tensor representing geometry of the LSF manifold M (G_{ij} is the trace-reversed Ricci tensor R_{ij} , which is itself the trace of the *Riemann curvature tensor* of the manifold M), while T_{ij} is the n -dimensional *stress-energy-momentum* tensor. This equation explicitly states that *psychophysics of the LSF is proportional to its geometry*. T_{ij} is important quantity, representing motivational *energy*, geometry-imposed *stress* and *momentum* of (loco)motion. As before, we have our ‘golden rule’: *the greater the T_{ij} -components, the higher the speed of cognitive processes and the lower the macroscopic fatigue*.

The choice between the geodesic action (6.148) and the Hilbert action (6.149) depends on our interpretation of time. If time is not included in the LSF manifold M (non-relativistic approach) then we choose the geodesic action. If time is included in the LSF manifold M (making it a relativistic-like n -dimensional space-time) then the Hilbert action is preferred. The first approach is more related to the information processing and the working memory. The later, space-time approach can be related to the long-term memory: we usually recall events closely associated with the times of their happening.

On the micro-level in the subspace LSF_{geom} we have the adaptive *sum over geometries*, represented by the path integral over all local (regional) Riemannian metrics $g_{ij} = g_{ij}(x)$ varying from point to point on M (modulo diffeomorphisms),

$$\langle Action | Intention \rangle_{geom} = \oint \mathcal{D}[wg_{ij}] e^{iS[g_{ij}]} \xrightarrow{Wick} \oint \mathcal{D}[wg_{ij}] e^{-S[g_{ij}]}, \quad (6.150)$$

where $\mathcal{D}[g_{ij}]$ is diffeomorphism equivalence class of $g_{ij}(x) \in M$.

To include the topological structure (e.g., a number of holes) in M , we can extend (6.150) as

$$\langle Action | Intention \rangle_{geom/top} = \sum_{\text{topol.}} \oint \mathcal{D}[wg_{ij}] e^{iS[g_{ij}]}, \quad (6.151)$$

where the topological sum is taken over all connectedness-components of M determined by the *Euler characteristic* χ of M . This type of integral defines the *theory of fluctuating geometries*, a propagator between $(n-1)$ -dimensional boundaries of the n -dimensional manifold M . One has to contribute a meaning to the integration over geometries. A key ingredient in doing so is to approximate (using simplicial approximation and Regge calculus [Misner *et al.* (1973)]) in a natural way the smooth structures of the manifold M by piecewise linear structures (mostly using topological simplices Δ). In this way, after the Wick-rotation (6.140), the integral (6.150–6.151) becomes a *simple statistical system*, given by partition function $Z = \sum_{\Delta} \frac{1}{C_{\Delta}} e^{-S_{\Delta}}$, where the summation is over all triangulations Δ of the manifold M , while C_T is the order of the automorphism group of the performed triangulation.

Micro-Level Geometry: the source of noise and stress in LSF.

The subspace LSF_{geom} is the source of noise, fluctuations and obstacles, as well as psycho-physical stress. Its micro-level is adaptive, reflecting the human ability to efficiently act within the noisy environment and under the stress conditions. By averaging it produces smooth geometry of certain curvature, which is at the same time the smooth psycho-physics. This macro-level geometry directly affects the memory fields and indirectly affects the (loco)motion transition paths.

The Mental Force Law. As an effective summary of this section, we state that the psychodynamic transition functor $\mathcal{TA} : INTENTION_{t_{ini}} \Rightarrow ACTION_{t_{fin}}$, defined by the generic path integral

(6.133), can be interpreted as a *mental force law*, analogous to our musculo-skeletal *covariant force law*, $F_i = mg_{ij}a^j$, and its associated *covariant force functor* $\mathcal{F}_* : TT^*M \rightarrow TTM$ [Ivancevic and Ivancevic (2006)].

6.5 Application: Witten's TQFT, SW-Monopoles and Strings

In this section we review three parts of modern physics associated to the name of the Fields Medalist Edward Witten: topological quantum field theory, Seiberg–Witten monopole theory and open superstring theory. They are all extensions of a general *quantum field theory* (QFT).

Recall that any QFT deals with smooth maps $\gamma : \Sigma \rightarrow M$ of Riemannian manifolds Σ and M such that the dimension of Σ is the dimension of the theory. On the set $Map(\Sigma, M)$ of all smooth maps $\gamma = \gamma(\phi)$, we also have defined an *action function* $S[\phi]$ of the *field variables* ϕ . A *non-relativistic* QFT studies *real-valued* (Euclidean) *path integrals* of the form

$$\int_{Map(\Sigma, M)} V(\phi) \mathcal{D}[\phi] e^{-S[\phi]/\hbar}, \quad (6.152)$$

where $\mathcal{D}[\phi]$ represents some measure on the space of paths, is the *Planck constant* and $V : Map(\Sigma, M) \rightarrow \mathbb{R}$ is an *insertion function*. The number $e^{-S[\phi]/\hbar}$ should be interpreted as the *probability amplitude* of the contribution of the map $\gamma : \Sigma \rightarrow M$ to the path integral. The associated integral

$$Z^E = \int_{Map(\Sigma, M)} d\phi e^{-S[\phi]/\hbar},$$

is the *partition function* of the theory. In a *relativistic* QFT, the space Σ has a Lorentzian metric of signature $(-, +, \dots, +)$. The first coordinate is reserved for time, the rest are for space. In this case, the real-valued path integral (6.152) is replaced with the *complex-valued path integral*

$$Z^M = \int_{Map(\Sigma, M)} V(\phi) \mathcal{D}[\phi] e^{iS[\phi]/\hbar}.$$

6.5.1 Topological Quantum Field Theory

Before we come to (super)strings, we give a brief on *topological quantum field theory* (TQFT), as developed by Ed Witten, from his original path integral point of view (see [Witten (1988a); Labastida and Lozano (1998)]).

TQFT originated in 1982, when Witten rewrote classical *Morse theory* (see section 3.10.5.1 above, as well as section 3.13.5.2 below) in Dick Feynman's language of quantum field theory [Witten (1982)]. Witten's arguments made use of Feynman's path integrals and consequently, at first, they were regarded as mathematically non-rigorous. However, a few years later, A. Floer reformulated a rigorous Morse–Witten theory [Floer (1987)] (that won a Fields medal for Witten). This trend in which some mathematical structure is first constructed by quantum field theory methods and then reformulated in a rigorous mathematical ground constitutes one of the tendencies in modern physics.

In TQFT our basic topological space is an n D Riemannian manifold M with a metric $g_{\mu\nu}$. Let us consider on it a set of fields $\{\phi_i\}$, and let $S[\phi_i]$ be a real functional of these fields which is regarded as the *action* of the theory. We consider ‘operators’, $O_\alpha(\phi_i)$, which are in general arbitrary functionals of the fields. In TQFT these functionals are real functionals labelled by some set of indices α carrying topological or group-theoretical data. The *vacuum expectation value* (VEV) of a product of these operators is defined as

$$\langle O_{\alpha_1} O_{\alpha_2} \cdots O_{\alpha_p} \rangle = \int [D\phi_i] O_{\alpha_1}(\phi_i) O_{\alpha_2}(\phi_i) \cdots O_{\alpha_p}(\phi_i) \exp(-S[\phi_i]).$$

A quantum field theory is considered *topological* if the following relation is satisfied:

$$\frac{\delta}{\delta g^{\mu\nu}} \langle O_{\alpha_1} O_{\alpha_2} \cdots O_{\alpha_p} \rangle = 0, \quad (6.153)$$

i.e., if the VEV of some set of selected operators is independent of the metric $g_{\mu\nu}$ on M . If such is the case those operators are called ‘observables’.

There are two ways to guarantee, at least formally, that condition (6.153) is satisfied. The first one corresponds to the situation in which both, the action $S[\phi_i]$, as well as the operators O_{α_i} are *metric independent*. These TQFTs are called of *Schwarz* type. The most important representative is *Chern–Simons gauge theory*. The second one corresponds to the case in which there exist a *symmetry*, whose infinitesimal form is denoted by δ , satisfying the following properties:

$$\delta O_{\alpha_i} = 0, \quad T_{\mu\nu} = \delta G_{\mu\nu}, \quad (6.154)$$

where $T_{\mu\nu}$ is the SEM-tensor of the theory, i.e.,

$$T_{\mu\nu}(\phi_i) = \frac{\delta}{\delta g^{\mu\nu}} S[\phi_i]. \quad (6.155)$$

The fact that δ in (6.154) is a symmetry of the theory implies that the transformations $\delta\phi_i$ of the fields are such that both $\delta A[\phi_i] = 0$ and $\delta O_{\alpha_i}(\phi_i) = 0$. Conditions (6.154) lead, at least formally, to the following relation for VEVs:

$$\begin{aligned} \frac{\delta}{\delta g^{\mu\nu}} \langle O_{\alpha_1} O_{\alpha_2} \cdots O_{\alpha_p} \rangle &= - \int [D\phi_i] O_{\alpha_1}(\phi_i) O_{\alpha_2}(\phi_i) \cdots O_{\alpha_p}(\phi_i) T_{\mu\nu} e^{-S[\phi_i]} \\ &= - \int [D\phi_i] \delta (O_{\alpha_1}(\phi_i) O_{\alpha_2}(\phi_i) \cdots O_{\alpha_p}(\phi_i) G_{\mu\nu} \exp(-S[\phi_i])) = 0, \end{aligned} \quad (6.156)$$

which implies that the quantum field theory can be regarded as topological. This second type of TQFTs are called of *Witten type*. One of its main representatives is the theory related to Donaldson invariants, which is a twisted version of $N = 2$ supersymmetric *Yang-Mills gauge theory*. It is important to remark that the symmetry δ must be a scalar symmetry, i.e., that its symmetry parameter must be a scalar. The reason is that, being a global symmetry, this parameter must be covariantly constant and for arbitrary manifolds this property, if it is satisfied at all, implies strong restrictions unless the parameter is a scalar.

Most of the TQFTs of *cohomological type* satisfy the relation:

$$S[\phi_i] = \delta \Lambda(\phi_i), \quad (6.157)$$

for some functional $\Lambda(\phi_i)$. This has far-reaching consequences, for it means that the topological observables of the theory, in particular the partition function, (path integral) itself are independent of the value of the coupling constant. Indeed, let us consider for example the VEV:

$$\langle O_{\alpha_1} O_{\alpha_2} \cdots O_{\alpha_p} \rangle = \int [D\phi_i] O_{\alpha_1}(\phi_i) O_{\alpha_2}(\phi_i) \cdots O_{\alpha_p}(\phi_i) e^{-\frac{1}{g^2} S[\phi_i]}. \quad (6.158)$$

Under a change in the coupling constant, $1/g^2 \rightarrow 1/g^2 - \Delta$, one has (assuming that the observables do not depend on the coupling), up to first-order in Δ :

$$\begin{aligned} \langle O_{\alpha_1} O_{\alpha_2} \cdots O_{\alpha_p} \rangle &\longrightarrow \langle O_{\alpha_1} O_{\alpha_2} \cdots O_{\alpha_p} \rangle \\ &+ \Delta \int [D\phi_i] \delta \left[O_{\alpha_1}(\phi_i) O_{\alpha_2}(\phi_i) \cdots O_{\alpha_p}(\phi_i) \Lambda(\phi_i) \exp \left(-\frac{1}{g^2} S[\phi_i] \right) \right] \\ &= \langle O_{\alpha_1} O_{\alpha_2} \cdots O_{\alpha_p} \rangle. \end{aligned} \quad (6.159)$$

Hence, observables can be computed either in the *weak coupling limit*, $g \rightarrow 0$, or in the *strong coupling limit*, $g \rightarrow \infty$.

So far we have presented a rather general definition of TQFT and made a series of elementary remarks. Now we will analyze some aspects of its structure. We begin pointing out that given a theory in which (6.154) holds one can build *correlators* which correspond to *topological invariants* (in the sense that they are invariant under deformations of the metric $g_{\mu\nu}$) just by considering the operators of the theory which are *invariant under the symmetry*. We will call these operators *observables*. In virtue of (6.156), if one of these operators can be written as a symmetry transformation of another operator, its presence in a correlation function will make it vanish. Thus we may identify operators satisfying (6.154) which differ by an operator which corresponds to a symmetry transformation of another operator. Let us denote the set of the resulting classes by $\{\Phi\}$. By restricting the analysis to the appropriate set of operators, one has that in fact,

$$\delta^2 = 0. \quad (6.160)$$

Property (6.160) has consequences on the features of TQFT. First, the *symmetry must be odd* which implies the presence in the theory of *commuting* and *anticommuting fields*. For example, the tensor $G_{\mu\nu}$ in (6.154) must be anticommuting. This is the first appearance of an odd non-spinorial field in TQFT. Those kinds of objects are standard features of *cohomological* TQFTs. Second, if we denote by Q the operator which implements this symmetry, the observables of the theory can be described as the cohomology classes of Q :

$$\{\Phi\} = \frac{\text{Ker } Q}{\text{Im } Q}, \quad Q^2 = 0. \quad (6.161)$$

Equation (6.154) means that in addition to the *Poincaré group* the theory possesses a symmetry generated by an odd version of the Poincaré group. The corresponding odd generators are constructed out of the tensor $G_{\mu\nu}$ in much the same way as the ordinary Poincaré generators are built out of $T_{\mu\nu}$. For example, if P_μ represents the ordinary momentum operator, there exists a corresponding odd one G_μ such that

$$P_\mu = \{Q, G_\mu\}. \quad (6.162)$$

Now, let us discuss the structure of the Hilbert space of the theory in virtue of the symmetries that we have just described. The states of this space must correspond to representations of the algebra generated by the

operators in the Poincaré groups and by Q . Furthermore, as follows from our analysis of operators leading to (6.161), if one is interested only in states $|\Psi\rangle$ leading to topological invariants one must consider states which satisfy

$$Q|\Psi\rangle = 0, \quad (6.163)$$

and two states which differ by a Q -exact state must be identified. The odd Poincaré group can be used to generate descendant states out of a state satisfying (6.163). The operators G_μ act non-trivially on the states and in fact, out of a state satisfying (6.163) we can build additional states using this generator. The simplest case consists of

$$\int_{\gamma_1} G_\mu |\Psi\rangle,$$

where γ_1 is a 1-cycle. One can verify using (6.154) that this new state satisfies (6.163):

$$Q \int_{\gamma_1} G_\mu |\Psi\rangle = \int_{\gamma_1} \{Q, G_\mu\} |\Psi\rangle = \int_{\gamma_1} P_\mu |\Psi\rangle = 0.$$

Similarly, one may construct other invariants tensoring n operators G_μ and integrating over n -cycles γ_n :

$$\int_{\gamma_n} G_{\mu_1} G_{\mu_2} \dots G_{\mu_n} |\Psi\rangle. \quad (6.164)$$

Notice that since the operator G_μ is odd and its algebra is Poincaré-like the integrand in this expression is an exterior differential n -form. These states also satisfy condition (6.163). Therefore, starting from a state $|\Psi\rangle \in \ker Q$ we have built a set of partners or descendants giving rise to a topological multiplet. The members of a multiplet have well defined *ghost* number. If one assigns ghost number -1 to the operator G_μ the state in (6.164) has ghost number $-n$ plus the ghost number of $|\Psi\rangle$. Now, n is bounded by the dimension of the manifold X . Among the states constructed in this way there may be many which are related via another state which is Q -exact, i.e., which can be written as Q acting on some other state. Let us try to single out representatives at each level of ghost number in a given topological multiplet.

Consider an $(n-1)$ -cycle which is the boundary of an n D surface,

$\gamma_{n-1} = \partial S_n$. If one builds a state taking such a cycle one finds ($P_\mu = -i\partial_\mu$),

$$\begin{aligned} \int_{\gamma_{n-1}} G_{\mu_1} G_{\mu_2} \dots G_{\mu_{n-1}} |\Psi\rangle &= i \int_{S_n} P_{[\mu_1} G_{\mu_2} G_{\mu_3} \dots G_{\mu_n]} |\Psi\rangle \quad (6.165) \\ &= iQ \int_{S_n} G_{\mu_1} G_{\mu_2} \dots G_{\mu_n} |\Psi\rangle, \end{aligned}$$

i.e., it is Q -exact. The square-bracketed subscripts in (6.165) denote that all indices between them must by antisymmetrized. In (6.165) use has been made of (6.162). This result tells us that the representatives we are looking for are built out of the homology cycles of the manifold X . Given a manifold X , the homology cycles are equivalence classes among cycles, the equivalence relation being that two n -cycles are equivalent if they differ by a cycle which is the boundary of an $n+1$ surface. Thus, knowledge on the homology of the manifold on which the TQFT is defined allows us to classify the representatives among the operators (6.164). Let us assume that X has dimension d and that its homology cycles are γ_{i_n} , ($i_n = 1, \dots, d_n$, $n = 0, \dots, d$), where d_n is the dimension of the n -homology group, and d the dimension of X . Then, the non-trivial partners or descendants of a given $|\Psi\rangle$ *highest-ghost-number state* are labelled in the following way:

$$\int_{\gamma_{i_n}} G_{\mu_1} G_{\mu_2} \dots G_{\mu_n} |\Psi\rangle, \quad (i_n = 1, \dots, d_n, \quad n = 0, \dots, d).$$

A similar construction to the one just described can be made for fields. Starting with a field $\phi(x)$ which satisfies,

$$[Q, \phi(x)] = 0, \quad (6.166)$$

one can construct other fields using the operators G_μ . These fields, which we call *partners* are antisymmetric tensors defined as,

$$\phi_{\mu_1 \mu_2 \dots \mu_n}^{(n)}(x) = \frac{1}{n!} [G_{\mu_1}, [G_{\mu_2} \dots [G_{\mu_n}, \phi(x)] \dots]], \quad (n = 1, \dots, d).$$

Using (6.162) and (6.166) one finds that these fields satisfy the so-called *topological descent equations*:

$$d\phi^{(n)} = i[Q, \phi^{(n+1)}],$$

where the subindices of the forms have been suppressed for simplicity, and the highest-ghost-number field $\phi(x)$ has been denoted as $\phi^{(0)}(x)$. These equations enclose all the relevant properties of the observables which are

constructed out of them. They constitute a very useful tool to build the observables of the theory.

6.5.2 *Seiberg–Witten Theory and TQFT*

Recall that the field of low-dimensional geometry and topology [Atiyah (1988b)] has undergone a dramatic phase of progress in the last decade of the 20th Century, prompted, to a large extent, by new ideas and discoveries in mathematical physics. The discovery of quantum groups [Drinfeld (1986)] in the study of the *Yang–Baxter equation* [Baxter (1982)] has reshaped the theory of knots and links [Jones (1985); Reshetikhin and Turaev (1991); Zhang *et. al.* (1991)]; the study of conformal field theory and quantum *Chern–Simons theory* [Witten (1989)] in physics had a profound impact on the theory of 3-manifolds; and most importantly, investigations of the classical Yang–Mills (YM) theory led to the creation of the *Donaldson theory* of 4-manifolds [Freed and Uhlenbeck (1984); Donaldson (1987)]. Witten [Witten (1994)] discovered a new set of invariants of 4-manifolds in the study of the Seiberg–Witten (SW) monopole equations, which have their origin in supersymmetric gauge theory. The SW theory, while closely related to Donaldson theory, is much easier to handle. Using SW theory, proofs of many theorems in Donaldson theory have been simplified, and several important new results have also been obtained [Taubes (1990); Taubes (1994)].

In [Zhang *et. al.* (1995)] a topological quantum field theory was introduced which reproduces the SW invariants of 4-manifolds. A geometrical interpretation of the 3D quantum field theory was also given.

6.5.2.1 *SW Invariants and Monopole Equations*

Recall that the SW monopole equations are classical field theoretical equations involving a $U(1)$ gauge field and a complex Weyl spinor on a 4D manifold. Let X denote the 4-manifold, which is assumed to be oriented and closed. If X is spin, there exist positive and negative spin bundles S^\pm of rank two. Introduce a complex line bundle $L \rightarrow X$. Let A be a connection on L and M be a section of the product bundle $S^+ \otimes L$. Recall that the SW monopole equations read

$$\mathcal{F}_{kl}^+ = -\frac{i}{2}\bar{M}\Gamma_{kl}M, \quad D_A M = 0, \quad (6.167)$$

where D_A is the twisted Dirac operator, $\Gamma_{ij} = \frac{1}{2}[\gamma_i, \gamma_j]$, and \mathcal{F}^+ represents the self-dual part of the curvature of L with connection A .

If X is not a spin manifold, then spin bundles do not exist. However, it is always possible to introduce the so called $Spin_c$ bundles $S^\pm \otimes L$, with L^2 being a line bundle. Then in this more general setting, the SW monopoles equations look formally the same as (6.167), but the M should be interpreted as a section of the the $Spin_c$ bundle $S^+ \otimes L$.

Denote by \mathcal{M} the moduli space of solutions of the SW monopole equations up to gauge transformations. Generically, this space is a manifold. Its virtual dimension is equal to the number of solutions of the following equations

$$\begin{aligned} (d\psi)_{kl}^+ + \frac{i}{2} (\bar{M}\Gamma_{kl}N + \bar{N}\Gamma_{kl}M) &= 0, & D_A N + \psi M &= 0, \\ \nabla_k \psi^k + \frac{i}{2} (\bar{N}M - \bar{M}N) &= 0, \end{aligned} \quad (6.168)$$

where A and M are a given solution of (6.167), $\psi \in \Omega^1(X)$ is a one form, $(d\psi)^+ \in \Omega^{2,+}(X)$ is the self dual part of the two form $d\psi$, and $N \in S^+ \otimes L$. The first two of the equations in (6.168) are the linearization of the monopole equations (6.167), while the last one is a gauge fixing condition. Though with a rather unusual form, it arises naturally from the dual operator governing gauge transformations

$$C : \Omega^0(X) \rightarrow \Omega^1(X) \oplus (S^+ \otimes L), \quad \phi \mapsto (-d\phi, i\phi M).$$

$$\text{Let } T : \Omega^1(X) \oplus (S^+ \otimes L) \rightarrow \Omega^0(X) \oplus \Omega^{2,+}(X) \oplus (S^- \otimes L),$$

be the operator governing equation (6.168), namely, the operator which allows us to rewrite (6.168) as $T(\psi, N) = 0$. Then T is an elliptic operator, the index $\text{Ind}(T)$ of which yields the virtual dimension of \mathcal{M} . A straightforward application of the Atiyah–Singer index Theorem gives

$$\text{Ind}(T) = -\frac{2\chi(X) + 3\sigma(X)}{4} + c_1(L)^2,$$

where $\chi(X)$ is the Euler character of X , $\sigma(X)$ its signature index and $c_1(L)^2$ is the square of the first Chern class of L evaluated on X in the standard way.

When $\text{Ind}(T)$ equals zero, the moduli space generically consists of a finite number of points, $\mathcal{M} = \{p_t : t = 1, 2, \dots, I\}$. Let ϵ_t denote the sign of the determinant of the operator T at p_t , which can be defined with

mathematical rigor. Then the SW invariant of the 4-manifold X is defined by $\sum_1^I \epsilon_t$.

The fact that this is indeed an invariant (i.e., independent of the metric) of X is not very difficult to prove, and we refer to [Witten (1994)] for details. As a matter of fact, the number of solutions of a system of equations weighted by the sign of the operator governing the equations (i.e., the analog of T) is a topological invariant in general [Witten (1994)]. This point of view has been extensively explored by Vafa and Witten [Vafa and Witten (1994)] within the framework of topological quantum field theory in connection with the so called S duality. Here we wish to explore the SW invariants following a similar line as that taken in [Witten (1988a); Vafa and Witten (1994)].

6.5.2.2 Topological Lagrangian

Introduce a *Lie super-algebra* with an odd generator Q and two even generators U and δ obeying the following (anti)commutation relations [Zhang *et. al.* (1995)]

$$[U, Q] = Q, \quad [Q, Q] = 2\delta, \quad [Q, \delta] = 0. \quad (6.169)$$

We will call U the ghost number operator, and Q the *BRST-operator*.

Let A be a connection of L and $M \in S^+ \otimes L$. We define the action of the super-algebra on these fields by requiring that δ coincide with a gauge transformation with a gauge parameter $\phi \in \Omega^0(X)$. The field multiplets associated with A and M furnishing representations of the super-algebra are (A, ψ, ϕ) , and (M, N) , where $\psi \in \Omega^1(X)$, $\phi \in \Omega^0(X)$, and N is a section of $S^+ \otimes L$. They transform under the action of the super-algebra according to

$$\begin{aligned} [Q, A_i] &= \psi_i, & [Q, M] &= N, \\ [Q, \psi_i] &= -\partial_i \phi, & [Q, N] &= i\phi M, & [Q, \phi] &= 0. \end{aligned}$$

We assume that both A and M have ghost number 0, and thus will be regarded as bosonic fields when we study their quantum field theory. The ghost numbers of other fields can be read off the above transformation rules. We have that ψ and N are of ghost number 1, thus are fermionic, and ϕ is of ghost number 2 and bosonic. Note that the multiplet (A, ψ, ϕ) is what one would get in the topological field theory for Donaldson invariants except that our gauge group is $U(1)$, while the existence of M and N is a new feature. Also note that both M and ψ have the wrong statistics.

In order to construct a quantum field theory which will reproduce the SW invariants as correlation functions, anti-ghosts and Lagrangian multipliers are also required. We introduce the anti-ghost multiplet $(\lambda, \eta) \in \Omega^0(X)$, such that

$$[U, \lambda] = -2\lambda, \quad [Q, \lambda] = \eta, \quad [Q, \eta] = 0,$$

and the Lagrangian multipliers $(\chi, H) \in \Omega^{2,+}(X)$, and $(\mu, \nu) \in S^- \otimes L$ such that

$$\begin{aligned} [U, \chi] &= -\chi, & [Q, \chi] &= H, & [Q, H] &= 0; \\ [U, \mu] &= -\mu, & [Q, \mu] &= \nu, & [Q, \nu] &= i\phi\mu. \end{aligned}$$

With the given fields, we construct the following functional which has ghost number -1:

$$\begin{aligned} V = \int_X \bigg\{ & [\nabla_k \psi^k + \frac{i}{2}(\overline{N}M - \overline{M}N)]\lambda - \chi^{kl} \left(H_{kl} - \mathcal{F}_{kl}^+ - \frac{i}{2}\overline{M}\Gamma_{kl}M \right) \\ & - \bar{\mu}(\nu - iD_A M) - \overline{(\nu - iD_A M)}\mu \bigg\}, \end{aligned} \quad (6.170)$$

where the indices of the tensorial fields are raised and lowered by a given metric g on X , and the integration measure is the standard $\sqrt{g}d^4x$. Also, \overline{M} and $\bar{\mu}$ etc. represent the Hermitian conjugate of the spinorial fields. In a formal language, $\overline{M} \in S^+ \otimes L^{-1}$ and $\bar{\mu}, \bar{\nu}, \overline{D_A M} \in S^- \otimes L^{-1}$. Following the standard procedure in constructing topological quantum field theory, we take the classical action of our theory to be [Zhang *et. al.* (1995)]: $S = [Q, V]$, which has ghost number 0. One can easily show that S is also BRST invariant, i.e., $[Q, S] = 0$, thus it is invariant under the full super-algebra (6.169).

The bosonic Lagrangian multiplier fields H and ν do not have any dynamics, and so can be eliminated from the action by using their equations of motion

$$H_{kl} = \frac{1}{2} \left(\mathcal{F}_{kl}^+ + \frac{i}{2}\overline{M}\Gamma_{kl}M \right), \quad \nu = \frac{1}{2}iD_A M. \quad (6.171)$$

Then we arrive at the following expression for the action [Zhang *et. al.*

(1995)]

$$\begin{aligned}
S = \int_X \bigg\{ & [-\Delta\phi + \overline{M}M\phi - i\overline{N}N]\lambda - [\nabla_k\psi^k + \frac{i}{2}(\overline{N}M - \overline{M}N)]\eta + 2i\phi\bar{\mu}\mu \\
& + (\overline{iD_A N - \gamma.\psi M})\mu - \bar{\mu}(iD_A N - \gamma.\psi M) \\
& - \chi^{kl} \left[(\nabla_k\psi^l - \nabla_l\psi^k)^+ + \frac{i}{2}(\overline{M}\Gamma_{kl}N + \overline{N}\Gamma_{kl}M) \right] \bigg\} + S_0,
\end{aligned} \tag{6.172}$$

where S_0 is given by

$$S_0 = \int_X \left\{ \frac{1}{4}|\mathcal{F}^+|^2 + \frac{i}{2}\overline{M}\Gamma M|^2 + \frac{1}{2}|D_A M|^2 \right\}.$$

It is interesting to observe that S_0 is nonnegative, and vanishes if and only if A and M satisfy the SW monopole equations. As pointed out in [Witten (1994)], S_0 can be rewritten as

$$S_0 = \int_X \left\{ \frac{1}{4}|\mathcal{F}^+|^2 + \frac{1}{4}|M|^4 + \frac{1}{8}R|M|^2 + g^{ij}\overline{D_i M}D_j M \right\},$$

where R is the scalar curvature of X associated with the metric g . If R is nonnegative over the entire X , then the only square integrable solution of the monopole equations (6.167) is A is a anti-self-dual connection and $M = 0$.

6.5.2.3 Quantum Field Theory

We will now investigate the quantum field theory defined by the classical action (6.172) with the path integral method. Let \mathcal{F} collectively denote all the fields. The partition function of the theory is defined by [Zhang *et. al.* (1995)]

$$Z = \int \mathcal{D}\mathcal{F} \exp(-\frac{1}{e^2}S),$$

where $e \in \mathbb{R}$ is the coupling constant. The integration measure $\mathcal{D}\mathcal{F}$ is defined on the space of all the fields. However, since S is invariant under the gauge transformations, we assume the integration over the gauge field to be performed over the gauge orbits of A . In other words, we fix a gauge for the A field using, say, a *Faddeev-Popov procedure*. This can be carried out in the standard manner, thus there is no need for us to spell out the details here. The integration measure $\mathcal{D}\mathcal{F}$ can be shown to be invariant under the super charge Q . Also, it does not explicitly involve the metric g of X .

Let W be any operator in the theory. Its correlation function is defined by

$$Z[W] = \int \mathcal{D}\mathcal{F} \exp(-\frac{1}{e^2}S)W.$$

It follows from the Q invariance of both the action S and the path integration measure that for any operator W ,

$$Z[[Q, W]] = \int \mathcal{D}\mathcal{F} \exp(-\frac{1}{e^2}S)[Q, W] = 0.$$

For the purpose of constructing topological invariants of the 4-manifold X , we are particularly interested in operators W which are BRST-closed,

$$[Q, W] = 0, \quad (6.173)$$

but not BRST-exact, i.e., can not be expressed as the (anti)-commutators of Q with other operators. For such a W , if its variation with respect to the metric g is *BRST* exact,

$$\delta_g W = [Q, W'], \quad (6.174)$$

then its correlation function $Z[W]$ is a topological invariant of X (by that we really mean that it does not depend on the metric g):

$$\delta_g Z[W] = \int \mathcal{D}\mathcal{F} \exp(-\frac{1}{e^2}S)[Q, W' - \frac{1}{e^2}\delta_g V.W] = 0.$$

In particular, the partition function Z itself is a topological invariant.

Another important property of the partition function is that it does not depend on the coupling constant e :

$$\frac{\partial Z}{\partial e^2} = \int \mathcal{D}\mathcal{F} \frac{1}{e^4} \exp(-\frac{1}{e^2}S)[Q, V] = 0.$$

Therefore, Z can be computed *exactly* in the limit when the coupling constant goes to zero. Such a computation can be carried out in the standard way: Let A^o, M^o be a solution of the equations of motion of A and M arising from the action S . We expand the fields A and M around this classical configuration,

$$A = A^o + ea, \quad M = M^o + em,$$

where a and m are the quantum fluctuations of A and M respectively. All the other fields do not acquire background components, thus are purely

quantum mechanical. We scale them by the coupling constant e , by setting N to eN , ϕ to $e\phi$ etc.. To the order $o(1)$ in e^2 , we have

$$Z = \sum_p \exp\left(-\frac{1}{e^2} S_{cl}^{(p)}\right) \int \mathcal{D}\mathcal{F}' \exp(-S_q^{(p)}),$$

where $S_q^{(p)}$ is the quadratic part of the action in the quantum fields and depends on the gauge orbit of the classical configuration A^o , M^o , which we label by p . Explicitly [Zhang *et. al.* (1995)],

$$\begin{aligned} S_q^{(p)} = & \int_X \left\{ [-\Delta\phi + \bar{M}^o M^o \phi - i\bar{N}N]\lambda - [\nabla_k \psi^k + \frac{i}{2}(\bar{N}M^o - \bar{M}^o N)]\eta + 2i\phi\bar{\mu}\mu \right. \\ & + \overline{(iD_{A^o}N - \gamma.\psi M^o)}\mu - \bar{\mu}(iD_{A^o}N - \gamma.\psi M^o) \\ & - \chi^{kl} \left[(\nabla_k \psi^l - \nabla_l \psi^k)^+ + \frac{i}{2}(\bar{M}^o \Gamma_{kl}N + \bar{N}\Gamma_{kl}M^o) \right] \\ & \left. + \frac{1}{4}|f^+ + \frac{i}{2}(\bar{m}\Gamma M^o + \bar{M}^o \Gamma m)|^2 + \frac{1}{2}|iD_{A^o}m + \gamma.aM^o|^2 \right\}, \end{aligned}$$

with f^+ the self-dual part of $f = da$. The classical part of the action is given by $S_{cl}^{(p)} = S_0|_{A=A^o, M=M^o}$. The integration measure $\mathcal{D}\mathcal{F}'$ has exactly the same form as $\mathcal{D}\mathcal{F}$ but with A replaced by a , and M by m , \bar{M} by \bar{m} respectively. Needless to say, the summation over p runs through all gauge classes of classical configurations.

Let us now examine further features of our quantum field theory. A gauge class of classical configurations may give a non-zero contribution to the partition function in the limit $e^2 \rightarrow 0$ only if $S_{cl}^{(p)}$ vanishes, and this happens if and only if A^o and M^o satisfy (6.167). Therefore, the SW monopole equations are recovered from the quantum field theory.

The equations of motion of the fields ψ and N in the semi-classical approximation can be easily derived from the quadratic action $S_q^{(p)}$, solutions of which are the zero modes of the quantum fields ψ and N . The equations of motion read

$$\begin{aligned} (d\psi)_{kl}^+ + \frac{i}{2}(\bar{M}^o \Gamma_{kl}N + \bar{N}\Gamma_{kl}M^o) &= 0, & D_{A^o}N + \gamma.\psi M^o &= 0, \\ \nabla_k \psi^k + \frac{i}{2}(\bar{N}M - \bar{M}N) &= 0. \end{aligned} \tag{6.175}$$

Note that they are exactly the same equations which we have already discussed in (6.168). The first two equations are the linearization of the monopole equations, while the last is a ‘gauge fixing condition’ for ψ . The

dimension of the space of solutions of these equations is the virtual dimension of the moduli space \mathcal{M} . Thus, within the context of our quantum field theoretical model, the virtual dimension of \mathcal{M} is identified with the number of the zero modes of the quantum fields ψ and N .

For simplicity we assume that there are no zero modes of ψ and N , i.e., the moduli space is zero-dimensional. Then no zero modes exist for the other two fermionic fields χ and μ . To compute the partition function in this case, we first observe that the quadratic action $S_q^{(p)}$ is invariant under the supersymmetry obtained by expanding Q to first order in the quantum fields around the monopole solution A^o , M^o (equations of motion for the nonpropagating fields H and ν should also be used.). This supersymmetry transforms the set of 8 real bosonic fields (each complex field is counted as two real ones; the a_i contribute 2 upon gauge fixing.) and the set of 16 fermionic fields to each other. Thus at a given monopole background we get [Zhang *et. al.* (1995)]

$$\int \mathcal{D}\mathcal{F}' \exp(-S_q^{(p)}) = \frac{\text{Pfaff}(\nabla_{\mathcal{F}})}{|\text{Pfaff}(\nabla_{\mathcal{F}})|} = \epsilon^{(p)},$$

where $\epsilon^{(p)}$ is +1 or -1. In the above equation, $\nabla_{\mathcal{F}}$ is the skew symmetric first order differential operator defining the fermionic part of the action $S_q^{(p)}$, which can be read off from $S_q^{(p)}$ to be $\nabla_{\mathcal{F}} = \begin{pmatrix} 0 & T \\ -T^* & 0 \end{pmatrix}$. Therefore, $\epsilon^{(p)}$ is the sign of the determinant of the elliptic operator T at the monopole background A^o , M^o , and the partition function $Z = \sum_p \epsilon^{(p)}$ coincides with the SW invariant of the 4-manifold X .

When the dimension of the moduli space \mathcal{M} is greater than zero, the partition function Z vanishes identically, due to integration over zero modes of the fermionic fields. In order to get any non trivial topological invariants for the underlying manifold X , we need to examine correlations functions of operators satisfying equations (6.173) and (6.174). A class of such operators can be constructed following the standard procedure [Witten (1994)]. We define the following set of operators

$$\begin{aligned}
 W_{k,0} &= \frac{\phi^k}{k!}, & W_{k,1} &= \psi W_{k-1,0}, \\
 W_{k,2} &= \mathcal{F} W_{k-1,0} - \frac{1}{2} \psi \wedge \psi W_{k-2,0}, \\
 W_{k,3} &= \mathcal{F} \wedge \psi W_{k-2,0} - \frac{1}{3!} \psi \wedge \psi \wedge \psi W_{k-3,0}, \\
 W_{k,4} &= \frac{1}{2} \mathcal{F} \wedge F W_{k-2,0} - \frac{1}{2} \mathcal{F} \wedge \psi \wedge \psi W_{k-3,0} - \frac{1}{4!} \psi \wedge \psi \wedge \psi \wedge \psi W_{k-4,0}.
 \end{aligned} \tag{6.176}$$

These operators are clearly independent of the metric g of X . Although they are not BRST invariant except for $W_{k,0}$, they obey the following equations [Zhang *et. al.* (1995)]

$$\begin{aligned}
 dW_{k,0} &= -[Q, W_{k,1}], & dW_{k,1} &= [Q, W_{k,2}], \\
 dW_{k,2} &= -[Q, W_{k,3}], & dW_{k,3} &= [Q, W_{k,4}], & dW_{k,4} &= 0,
 \end{aligned}$$

which allow us to construct BRST invariant operators from the the W 's in the following way: Let X_i , $i = 1, 2, 3$, $X_4 = X$, be compact manifolds without boundary embedded in X . We assume that these submanifolds are homologically nontrivial. Define

$$\widehat{O}_{k,0} = W_{k,0}, \quad \widehat{O}_{k,i} = \int_{X_i} W_{k,i}, \quad (i = 1, 2, 3, 4). \tag{6.177}$$

As we have already pointed out, $\widehat{O}_{k,0}$ is BRST invariant. It follows from the descendent equations that

$$[Q, \widehat{O}_{k,i}] = \int_{X_i} [Q, W_{k,i}] = \int_{X_i} dW_{k,i-1} = 0.$$

Therefore the operators \widehat{O} indeed have the properties (6.173) and (6.174). Also, for the boundary ∂K of an $i + 1$ D manifold K embedded in X , we have

$$\int_{\partial K} W_{k,i} = \int_K dW_{k,i} = [Q, \int_K W_{k,i+1}],$$

is BRST trivial. The correlation function of $\int_{\partial K} W_{k,i}$ with any BRST invariant operator is identically zero. This in particular shows that the \widehat{O} 's only depend on the homological classes of the submanifolds X_i .

6.5.2.4 Dimensional Reduction and 3D Field Theory

In this subsection we dimensionally reduce the quantum field theoretical model for the SW invariant from 4D to 3D, thus to get a new topological quantum field theory defined on 3-manifolds. Its partition function yields a 3-manifold invariant, which can be regarded as the SW version of Casson's invariant [Akbulut and McCarthy (1990); Taubes (1994)].

We take the 4-manifold X to be of the form $Y \times [0, 1]$ with Y being a compact 3-manifold without boundary. The metric on X will be taken to be

$$(ds)^2 = (dt)^2 + g_{ij}(x)dx^i dx^j,$$

where the 'time' t -independent $g(x)$ is the Riemannian metric on Y . We assume that Y admits a spin structure which is compatible with the $Spin_c$ structure of X , i.e., if we think of Y as embedded in X , then this embedding induces maps from the $Spin_c$ bundles $S^\pm \otimes L$ of X to $\tilde{S} \otimes L$, where \tilde{S} is a spin bundle and L is a line bundle over Y .

To perform the dimensional reduction, we impose the condition that all fields are t in dependent. This leads to the following action [Zhang *et. al.* (1995)]

$$\begin{aligned} S = \int \sqrt{g} d^3x \bigg\{ & [-\Delta\phi + \bar{M}M\phi - i\bar{N}N]\lambda - [\nabla_k\psi^k + \frac{i}{2}(\bar{N}M - \bar{M}N)]\eta + 2i\phi\bar{\mu}\mu \\ & + [\bar{i}(D_A + b)N - (\sigma\cdot\psi - \tau)\bar{M}]\mu - \bar{\mu}[i(D_A + b)N - (\sigma\cdot\psi - \tau)M] \\ & - 2\chi^k [-\partial_k\tau + *(\nabla\psi)_k - \bar{M}\sigma_k N - \bar{N}\sigma_k M] \\ & + \frac{1}{4}|*\mathcal{F} - \partial b - \bar{M}\sigma M|^2 + \frac{1}{2}|(D_A + b)M|^2 \bigg\}, \end{aligned} \quad (6.178)$$

where the k is a 3D index, and σ_k are the Pauli matrices. The fields $b, \tau \in \Omega^0(Y)$ respectively arose from A_0 and ψ_0 of the 4D theory, while the meanings of the other fields are clear. The BRST symmetry in 4D carries over to the 3D theory. The BRST transformations rules for (A_i, ψ_i, ϕ) , $i = 1, 2, 3$, (M, N) , and (λ, η) are the same as before, but for the other fields, we have

$$\begin{aligned} [Q, b] &= \tau, & [Q, \tau] &= 0, \\ [Q, \chi_k] &= \frac{1}{2}(*\mathcal{F}_k - \partial_k b - \bar{M}\sigma_k M), \\ [Q, \mu] &= \frac{1}{2}i(D_A + b)M. \end{aligned}$$

The action S is cohomological in the sense that $S = [Q, V_3]$, with V_3 being the dimensionally reduced version of V defined by (6.170), and $[Q, S] = 0$. Thus it gives rise to a topological field theory upon quantization. The partition function of the theory

$$Z = \int \mathcal{D}\mathcal{F} \exp(-\frac{1}{e^2}S),$$

can be computed exactly in the limit $e^2 \rightarrow 0$, as it is coupling constant independent. We have, as before,

$$Z = \sum_p \exp(-\frac{1}{e^2}S_{cl}^{(p)}) \int \mathcal{D}\mathcal{F}' \exp(-S_q^{(p)}),$$

where $S_q^{(p)}$ is the quadratic part of S expanded around a classical configuration with the classical parts for the fields A, M, b being A^o, M^o, b^o , while those for all the other fields being zero. The classical action $S_{cl}^{(p)}$ is given by

$$S_{cl}^{(p)} = \int_Y \left\{ \frac{1}{4} |* \mathcal{F}^o - db^o - \bar{M}^o \sigma M^o|^2 + \frac{1}{2} |(D_{A^o} + b^o)M^o|^2 \right\},$$

which can be rewritten as [Zhang *et. al.* (1995)]

$$S_{cl}^{(p)} = \int_Y \left\{ \frac{1}{4} |* \mathcal{F}^o - \bar{M}^o \sigma M^o|^2 + \frac{1}{2} |D_{A^o} M^o|^2 + \frac{1}{2} |db^o|^2 + \frac{1}{2} |b^o M^o|^2 \right\}.$$

In order for the classical configuration to have non-vanishing contributions to the partition function, all the terms in $S_{cl}^{(p)}$ should vanish separately. Therefore,

$$* \mathcal{F}^o - \bar{M}^o \sigma M^o = 0, \quad D_{A^o} M^o = 0, \quad b^o = 0, \quad (6.179)$$

where the last condition requires some explanation. When we have a trivial solution of the equations (6.179), it can be replaced by the less stringent condition $db^o = 0$. However, in a more rigorous treatment of the problem at hand, we in general perturb the equations (6.179), then the trivial solution does not arise.

Let us define an operator

$$\begin{aligned} \tilde{T} : \Omega^0(Y) \oplus \Omega^1(Y) \oplus (\tilde{S} \otimes L) &\rightarrow \Omega^0(Y) \oplus \Omega^1(Y) \oplus (\tilde{S} \otimes L), \\ (\tau, \psi, N) &\mapsto (-d^* \psi + \frac{i}{2} (\bar{N} M - \overline{M} N), \quad *(d\psi) - d\tau - \bar{N} \sigma M - \overline{M} \sigma N, \\ &\quad iD_A N - (\sigma.\psi - \tau)M), \end{aligned} \quad (6.180)$$

where the complex bundle $\tilde{S} \otimes L$ should be regarded as a real one with twice the rank. This operator is self-adjoint, and is also obviously elliptic. We will assume that it is Fredholm as well. In terms of \tilde{T} , the equations of motion of the fields χ^i and μ can be expressed as [Zhang *et. al.* (1995)] $\tilde{T}^{(p)}(\tau, \psi, N) = 0$, where $\tilde{T}^{(p)}$ is the operator \tilde{T} with the background fields (A^o, M^o) belonging to the gauge class p of classical configurations.

When the kernel of \tilde{T} is zero, the partition function Z does not vanish identically. An easy computation leads to $Z = \sum_p \epsilon^{(p)}$, where the sum is over all gauge inequivalent solutions of (6.179), and $\epsilon^{(p)}$ is the sign of the determinant of $\tilde{T}^{(p)}$.

A rigorous definition of the sign of the $\det(\tilde{T})$ can be devised. However, if we are to compute only the absolute value of Z , then it is sufficient to know the sign of $\det(\tilde{T})$ relative to a fixed gauge class of classical configurations. This can be achieved using the *mod* -2 spectral flow of a family of Fredholm operators \tilde{T}_t along a path of solutions of (6.179). More explicitly, let (A^o, M^o) belong to the gauge class of classical configurations p , and $(\tilde{A}^o, \tilde{M}^o)$ in \tilde{p} . We consider the solution of the SW equation on $X = Y \times [0, 1]$ with $A_0 = 0$ and also satisfying the following conditions

$$(A, M)|_{t=0} = (A^o, M^o), \quad (A, M)|_{t=1} = (\tilde{A}^o, \tilde{M}^o).$$

Using this solution in \tilde{T} results in a family of *Fredholm operators*, which has zero kernels at $t = 0$ and 1 . The spectral flow of \tilde{T}_t , denoted by $q(p, \tilde{p})$, is defined to be the number of eigenvalues which cross zero with a positive slope minus the number which cross zero with a negative slope. This number is a well defined quantity, and is given by the index of the operator $\frac{\partial}{\partial t} - \tilde{T}_t$. In terms of the spectral flow, we have [Zhang *et. al.* (1995)]

$$\frac{\det(\tilde{T}^{(p)})}{\det(\tilde{T}^{(\tilde{p})})} = (-1)^{q(p, \tilde{p})}.$$

Equations (6.179) can be derived from the functional

$$S_{c-s} = \frac{1}{2} \int_Y A \wedge \mathcal{F} + i \int_Y \sqrt{g} d^3 x \bar{M} D_A M.$$

(It is interesting to observe that this is almost the standard Lagrangian of a $U(1)$ Chern–Simons theory coupled to spinors, except that we have taken M to have bosonic statistics.) S_{c-s} is gauge invariant modulo a constant arising from the Chern–Simons term upon a gauge transformation.

Therefore, $(\frac{\delta S_{c-s}}{\delta A}, \frac{\delta S_{c-s}}{\delta M})$ defines a vector field on the quotient space of all $U(1)$ connections \mathcal{A} tensored with the $\tilde{S} \times L$ sections by the $U(1)$ gauge group G , i.e., $\mathcal{W} = (\mathcal{A} \times (\tilde{S} \otimes L))/G$. Solutions of (6.179) are zeros of this vector field, and $\tilde{T}^{(p)}$ is the Hessian at the point $p \in \mathcal{W}$. Thus the partition Z is nothing else but the Euler character of \mathcal{W} . This geometrical interpretation will be spelt out more explicitly in the next subsection by re-interpreting the theory using the *Mathai–Quillen formula* [Mathai and Quillen (1986)].

6.5.2.5 Geometrical Interpretation

To elucidate the geometric meaning of the 3D theory obtained in the last section, we now cast it into the framework of Atiyah and Jeffrey [Atiyah and Jeffrey (1990)]. Let us briefly recall the geometric set up of the *Mathai–Quillen formula* as reformulated in reference [Atiyah and Jeffrey (1990)]. Let P be a Riemannian manifold of dimension $2m + \dim G$, and G be a compact Lie group acting on P by isometries. Then $P \rightarrow P/G$ is a principle bundle. Let V be a $2m$ dimensional real vector space, which furnishes a representation $G \rightarrow SO(2m)$. Form the associated vector bundle $P \times_G V$. Now the Thom form of $P \times_G V$ can be expressed [Zhang *et. al.* (1995)]

$$U = \frac{\exp(-x^2)}{(2\pi)^{\dim G \pi^m}} \int \exp \left\{ \frac{i\chi\phi\chi}{4} + i\chi dx - i\langle \delta\nu, \lambda \rangle - \langle \phi, R\lambda \rangle + \langle \nu, \eta \rangle \right\} \mathcal{D}\eta \mathcal{D}\chi \mathcal{D}\phi \mathcal{D}\lambda, \quad (6.181)$$

where $x = (x^1, \dots, x^{2m})$ is the coordinates of V , ϕ and λ are bosonic variables in the Lie algebra \mathfrak{g} of G , and η and χ are Grassmannian variables valued in the Lie algebra and the tangent space of the fiber respectively. In the above equation, C maps any $\eta \in \mathfrak{g}$ to the element of the vertical part of TP generated by η ; ν is the \mathfrak{g} -valued one form on P defined by $\langle \nu(\alpha), \eta \rangle = \langle \alpha, C(\eta) \rangle$, for all vector fields α ; and $R = C^*C$. Also, δ is the exterior derivative on P .

Now we choose a G invariant map $s : P \rightarrow V$, and pull back the Thom form U . Then the top form on P in s^*U is the Euler class. If $\{\delta p\}$ forms a basis of the cotangent space of P (note that ν and δs are one forms on P), we replace it by a set of Grassmannian variables $\{\psi\}$ in s^*U , then integrate

them away. We arrive at

$$\Upsilon = \frac{1}{(2\pi)^{\dim G} \pi^m} \int \exp \left\{ -|s|^2 + \frac{i\chi\phi\chi}{4} + i\chi\delta s - i\langle\delta\nu, \lambda\rangle \right. \\ \left. - \langle\phi, R\lambda\rangle + \langle\psi, C\eta\rangle \right\} \mathcal{D}\eta \mathcal{D}\chi \mathcal{D}\phi \mathcal{D}\lambda \mathcal{D}\psi, \quad (6.182)$$

the precise relationship of which with the Euler character of $P \times_G V$ is

$$\int_P \Upsilon = \text{Vol}(G) \chi(P \times_G V).$$

It is rather obvious that the action S defined by (6.172) for the 4D theory can be interpreted as the exponent in the integrand of (6.182), if we identify P with $\mathcal{A} \times \Gamma(W^+)$, and V with $\Omega^{2,+}(X) \times \Gamma(W^-)$, and set $s = (\mathcal{F}^+ + \frac{i}{2} \bar{M} \Gamma M, D_A M)$. Here \mathcal{A} is the space of all $U(1)$ connections of $\det(W^+)$, and $\Gamma(W^\pm)$ are the sections of $S^\pm \otimes L$ respectively.

For the 3D theory, we wish to show that the partition function yields the Euler number of \mathcal{W} . However, the tangent bundle of \mathcal{W} cannot be regarded as an associated bundle with the principal bundle, for which for the formulae (6.181) or (6.182) can readily apply, some further work is required.

Let P be the principal bundle over P/G , V, V' be two orthogonal representations of G . Suppose there is an embedding from $P \times_G V' \rightarrow P \times_G V$ via a G -map $\gamma(p) : V' \rightarrow V$ for $p \in P$. Denote the resulting quotient bundle as E . In order to derive the Thom class for E , one needs to choose a section of E , or equivalently, a G -map $s : P \rightarrow V$ such that $s(p) \in (\text{Im} \gamma(p))^\perp$. Then the Euler class of E can be expressed as $\pi_* \rho^* U$, where U is the Thom class of $P \times_G V$, ρ is a G -map: $P \times V' \rightarrow P \times V$ defined by

$$\rho(p, \tau) = (p, \gamma(p)\tau + s(p)),$$

and π_* is the integration along the fiber for the projection $\pi : P \times V' \rightarrow P/G$. Explicitly, [Zhang *et. al.* (1995)]

$$\pi_* \rho^*(U) = \int \exp \left\{ -|\gamma(p)\tau + s(p)|^2 + i\chi\phi\chi + i\chi\delta(\gamma(p)\tau + s(p)) \right. \\ \left. - i\langle\delta\nu, \lambda\rangle - \langle\phi, R\lambda\rangle + \langle\nu, C\eta\rangle \right\} \mathcal{D}\chi \mathcal{D}\phi \mathcal{D}\tau \mathcal{D}\eta \mathcal{D}\lambda. \quad (6.183)$$

Consider the exact sequence

$$0 \longrightarrow (\mathcal{A} \times \Gamma(W)) \times_G \Omega^0(Y) \xrightarrow{j} (\mathcal{A} \times \Gamma(W)) \times_G (\Omega^1(Y) \times \Gamma(W)),$$

where $j_{(A,M)} : b \mapsto (-db, bM)$ (assuming that $M \neq 0$). Then the tangent bundle of $\mathcal{A} \times_G \Gamma(W)$ can be Regarded as the quotient bundle

$$(\mathcal{A} \times \Gamma(W)) \times_G (\Omega^1(Y) \times \Gamma(W)) / \text{Im}(j).$$

We define a vector field on $\mathcal{A} \times_G \Gamma(W)$ by

$$s(A, M) = (*\mathcal{F}_A - \bar{M}\sigma M, D_A M),$$

which lies in $\text{Im}(j)^\perp$:

$$\int_Y (*\mathcal{F}_A - \bar{M}\sigma M) \wedge *(-db) + \int_Y \sqrt{g} d^3x \langle D_A M, bM \rangle = 0, \quad (6.184)$$

where we have used the short hand notation $\langle M_1, M_2 \rangle = \frac{1}{2}(\bar{M}_1 M_2 + \bar{M}_2 M_1)$.

Formally applying the formula (6.183) to the present infinite-dimensional situation, we get the *Euler class* $\pi_* \rho^*(U)$ for the tangent bundle $T(\mathcal{A} \times_G \Gamma(W))$, where ρ is the \mathcal{G} -invariant map ρ is defined by

$$\rho : \Omega^0(Y) \longrightarrow \Omega^1(Y) \times \Gamma(W), \quad \rho(b) = (-db + *\mathcal{F}_A - \bar{M}\sigma M, (D_A + b)M),$$

π is the projection $(\mathcal{A} \times \Gamma(W)) \times_G \Omega^0(Y) \longrightarrow \mathcal{A} \times_G \Gamma(W)$, and π_* signifies the integration along the fiber. Also U is the Thom form of the bundle

$$(\mathcal{A} \times \Gamma(W)) \times_G (\Omega^1(Y) \times \Gamma(W)) \longrightarrow \mathcal{A} \times_G \Gamma(W).$$

To get a concrete feel about U , we need to explain the geometry of this bundle. The metric on Y and the Hermitian metric $\langle \cdot, \cdot \rangle$ on $\Gamma(W)$ naturally define a connection. The *Maurer-Cartan connection* on $\mathcal{A} \longrightarrow \mathcal{A}/G$ is flat while the Hermitian connection on has the curvature $i\phi\mu \wedge \bar{\mu}$. This gives the expression of term $i(\chi, \mu)\phi(\chi, \mu)$ in (6.182) in our case.

In our infinite-dimensional setting, the map C is given by

$$C : \Omega^0(Y) \longrightarrow T_{(A,M)}(\mathcal{A} \times \Gamma(W)), \quad C(\eta) = (-d\eta, i\eta M),$$

and its dual is given by

$$C^* : \Omega^1(Y) \times \Gamma(W) \longrightarrow \Omega^0(Y), \quad C^*(\psi, N) = -d^*\psi + \langle N, iM \rangle.$$

The one form $\langle \nu, \eta \rangle$ on $\mathcal{A} \times \Gamma(W)$ takes the value

$$\langle (\psi, N), C\eta \rangle = \langle -d^*\psi, \eta \rangle + \langle N, iM \rangle \eta$$

on the vector field (ψ, N) . We also easily get $R(\lambda) = -\Delta\lambda + \langle M, M \rangle \lambda$, where $\Delta = d^*d$. The $\langle \delta\nu, \lambda \rangle$ is a two form on $\mathcal{A} \times \Gamma(W)$ whose value on $(\psi_1, N_1), (\psi_2, N_2)$ is $-\langle N_1, N_2 \rangle \lambda$.

Combining all the information together, we arrive at the following formula,

$$\begin{aligned} \pi_* \rho^*(U) = & \int \exp \left\{ -\frac{1}{2} |\rho|^2 + i(\chi, \mu) \delta\rho + 2i\phi\mu\bar{\mu} \right. \\ & + \langle \Delta\phi, \lambda \rangle - \phi\lambda\langle M, M \rangle + i\langle N, N \rangle \lambda \\ & \left. + \langle \nu, \eta \rangle \right\} \mathcal{D}\chi \mathcal{D}\phi \mathcal{D}\lambda \mathcal{D}\eta \mathcal{D}b. \end{aligned} \quad (6.185)$$

Note that the 1-form $i(\chi, \mu) \delta\rho$ on $\mathcal{A} \times \Gamma(W) \times \Omega^0(Y)$ contacted with the vector field (ϕ, N, b) leads to

$$2\chi^k \left[-\partial_k \tau + *(\nabla\psi)_k - \bar{M}\sigma_k N - \bar{N}\sigma_k M \right] + 2\langle \mu, [i(D_A + b)N - (\sigma.\psi - \tau)M] \rangle;$$

and the relation (6.184) gives $|\rho|^2 = |*\mathcal{F} - \bar{M}\sigma M|^2 + |db|^2 + |D_A M|^2 + b^2|M|^2$. Finally we get the *Euler character*

$$\pi_* \rho^*(U) = \int \exp(-S) \mathcal{D}\chi \mathcal{D}\phi \mathcal{D}\lambda \mathcal{D}\eta \mathcal{D}b, \quad (6.186)$$

where S is the action (6.178) of the 3D theory defined on the manifold Y .

Integrating (6.186) over $\mathcal{A} \times_G \Gamma(W)$ leads to the Euler number

$$\sum_{[(A, M)]: s(A, M)=0} \epsilon^{(A, M)},$$

which coincides with the partition function Z of our 3D theory.

6.5.3 TQFTs Associated with SW-Monopoles

Recall that TQFTs are often used to study topological nature of manifolds. In particular, 3D and 4D TQFTs are well developed. The most well-known 3D TQFT would be the Chern–Simons theory, whose partition function gives *Ray–Singer torsion* of 3-manifolds and the other topological invariants can be obtained as gauge invariant observables i.e., *Wilson loops*. The correlation functions can be identified with knot or link invariants e.g., Jones polynomial or its generalizations. On the other hand, in 4D, a twisted $N = 2$ supersymmetric YM theory developed by Witten [Witten (1988a)] also has a nature of TQFT. This YM theory can be interpreted as *Donaldson theory* and the *correlation functions* are identified with *Donaldson polynomials*, which classify smooth structures of

topological 4-manifolds. A new TQFT on 4-manifolds was discovered in SW studies of electric-magnetic duality of supersymmetric gauge theory. As discussed before, Seiberg and Witten [Seiberg and Witten (1994a); Seiberg and Witten (1994b)] studied the electric-magnetic duality of $N = 2$ supersymmetric $SU(2)$ YM gauge theory, by using a version of *Montonen-Olive duality* and obtained exact solutions. According to this result, the exact *low energy effective action* can be determined by a certain elliptic curve with a parameter $u = \langle \text{Tr}(\phi)^2 \rangle$, where ϕ is a complex scalar field in the adjoint representation of the gauge group, describing the quantum moduli space. For large u , the theory is weakly-coupled and semi-classical, but at $u = \pm\Lambda^2$ corresponding to strong coupling regime, where Λ is the dynamically generated mass scale, the elliptic curve becomes singular and the situation of the theory changes drastically. At these singular points, magnetically charged particles become massless. Witten showed that at $u = \pm\Lambda^2$ the TQFT was related to the moduli problem of counting the solution of the (Abelian) ‘Seiberg-Witten monopole equations’ [Witten (1994)] and it gave a dual description for the $SU(2)$ Donaldson theory.

It turns out that in 3D a particular TQFT of *Bogomol’nyi monopoles* can be obtained from a dimensional reduction of Donaldson theory and the partition function of this theory gives the so-called *Casson invariant* of 3-manifolds [Atiyah and Jeffrey (1990)].

Ohta [Ohta (1998)] discussed TQFTs associated with the 3D version of both Abelian and non-Abelian SW-monopoles, by applying *Batalin-Vilkovisky quantization* procedure. In particular, Ohta constructed the topological actions, topological observables and BRST transformation rules.

In this subsection, mainly following [Ohta (1998)], we will discuss TQFTs associated with both Abelian and non-Abelian SW-monopoles. We will use the following notation.

Let X be a compact orientable Spin 4-manifold without boundary and $g_{\mu\nu}$ be its Riemannian metric tensor (with $g = \det g_{\mu\nu}$). Here we use x_μ as the local coordinates on X . γ_μ are Dirac’s gamma matrices and $\sigma_{\mu\nu} = [\gamma_\mu, \gamma_\nu]/2$ with $\{\gamma_\mu, \gamma_\nu\} = g_{\mu\nu}$. M is a Weyl fermion and \bar{M} is a complex conjugate of M . (We will suppress spinor indices.) The Lie algebra \mathfrak{g} is defined by $[T^a, T^b] = if_{abc}T^c$, where T^a is a generator normalized as $\text{Tr}(T^a T^b) = \delta^{ab}$. The symbol f_{abc} is a structure constant of \mathfrak{g} and is antisymmetric in its indices. The Greek indices μ, ν, α etc run from 0 to 3. The Roman indices a, b, c, \dots are used for the Lie algebra indices running from 1 to $\dim \mathfrak{g}$, whereas i, j, k, \dots are the indices for space coordinates. Space-time indices are raised and lowered with $g_{\mu\nu}$. The repeated indices

are assumed to be summed. $\epsilon_{\mu\nu\rho\sigma}$ is an antisymmetric tensor with $\epsilon_{0123} = 1$. We often use the abbreviation of roman indices as $\theta = \theta^a T^a$ etc., in order to suppress the summation over Lie algebra indices.

Brief Review of TQFT

Firstly, we give a brief review of TQFT (compare with Witten's TQFT presented in subsection 6.5.1 above).

Let ϕ be any field content. For a local symmetry of ϕ , we can construct a nilpotent *BRST-operator* Q_B ($Q_B^2 = 0$). The variation of any functional \mathcal{O} of ϕ is denoted by $\delta\mathcal{O} = \{Q_B, \mathcal{O}\}$, where the bracket $\{*, *\}$ represents a graded commutator, that is, if \mathcal{O} is bosonic, the bracket means a commutator $[*, *]$ and otherwise it is an anti-bracket. Now, we can give the definition of topological field theory, as given in [Birmingham *et. al.* (1991)]:

A topological field theory consists of:

- (1) *a collection of Grassmann graded fields ϕ on an nD Riemannian manifold X with a metric g ,*
- (2) *a nilpotent Grassmann odd operator Q ,*
- (3) *physical states to be Q -cohomology classes,*
- (4) *an energy-momentum tensor $T_{\alpha\beta}$ which is Q -exact for some functional $V_{\alpha\beta}$ such as*

$$T_{\alpha\beta} = \{Q, V_{\alpha\beta}(\phi, g)\}.$$

In this definition, Q is often identified with Q_B and is in general independent of the metric. Now, recall that there are two broad types of TQFTs satisfying this definition and they are classified into Witten-type [Witten (1994)] or Schwarz-type [Schwarz ((1978))].

For *Witten-type* TQFT, the quantum action S_q which comprises the classical action, ghost and gauge fixing terms, can be represented by $S_q = \{Q_B, V\}$, for some function V of metric and fields and BRST charge Q_B . Under the metric variation δ_g of the partition function Z , it is easy to see that

$$\begin{aligned} \delta_g Z &= \int \mathcal{D}\phi \, e^{-S_q} \left(-\frac{1}{2} \int_X d^n x \sqrt{g} \delta g^{\alpha\beta} T_{\alpha\beta} \right) \\ &= \int \mathcal{D}\phi \, e^{-S_q} \{Q, \chi\} \equiv \langle \{Q, \chi\} \rangle = 0, \end{aligned} \quad (6.187)$$

$$\text{where} \quad \chi = -\frac{1}{2} \int_X d^n x \sqrt{g} \delta g^{\alpha\beta} V_{\alpha\beta}.$$

The last equality in (6.187) follows from the BRST invariance of the vacuum and means that Z is independent of the local structure of X , that is, Z is a *topological invariant* of X .

In general, for Witten type theory, Q_B can be constructed by an introduction of a topological shift with other local gauge symmetry [Ohta (1998)]. For example, in order to get the topological YM theory on four manifold M^4 , we introduce the shift in the gauge transformation for the gauge field A_μ^a such as $\delta A_\mu^a = D_\mu \theta^a + \epsilon_\mu^a$, where D_μ is a covariant derivative, θ^a and ϵ_μ^a are the (Lie algebra valued) usual gauge transformation parameter and topological shift parameter, respectively. In order to see the role of this shift, let us consider the first *Pontryagin class* on M^4 given by

$$S = \frac{1}{8} \int_{M^4} \epsilon^{\mu\nu\rho\sigma} F_{\mu\nu}^a F_{\rho\sigma}^a d^4x, \quad (6.188)$$

where $F_{\mu\nu}^a$ is a field strength of the gauge field. We can easily check the invariance of (6.188) under the action of δ . In this sense, (6.188) has a larger symmetry than the usual YM gauge symmetry. Taking this into account, we can construct the topological YM gauge theory. We can also consider similar ‘topological’ shifts for matter fields.

In addition, in general, Witten type topological field theory can be obtained from the quantization of some Langevin equations. This approach has been used for the construction of several topological field theories, e.g., supersymmetric quantum mechanics, topological sigma models or Donaldson theory (see [Birmingham *et. al.* (1991)]).

On the other hand, *Schwarz-type TQFT* [Schwarz ((1978))] begins with any metric independent classical action S_c as a starting point, but S_c is assumed not to be a total derivative. Then the quantum action (up to gauge fixing term) can be written by

$$S_q = S_c + \{Q, V(\phi, g)\}, \quad (6.189)$$

for some function V . For this quantum action, we can easily check the topological nature of the partition function, but note that the energy-momentum tensor contributes only from the second term in (6.189). One of the differences between Witten type and Schwarz type theories can be seen in this point. Namely, the energy-momentum tensor of the classical action for Schwarz type theory vanishes because it is derived as a result of metric variation.

Finally, we comment on the local symmetry of Schwarz type theory. Let

us consider the Chern–Simons theory as an example. The classical action,

$$S_{CS} = \int_{M^3} d^3x \left(A \wedge dA + \frac{2}{3} A \wedge A \wedge A \right), \quad (6.190)$$

is a *topological invariant*, which gives the *second Chern class* of 3-manifold M^3 . As is easy to find, S_{CS} is not invariant under the topological gauge transformation, although it is YM gauge invariant. Therefore the quantization is proceeded by the standard BRST method. This is a general feature of Schwarz-type TQFT.

6.5.3.1 Dimensional Reduction

First, let us recall the SW monopole equations in 4D. We assume that X has Spin structure. Then there exist rank two positive and negative spinor bundles S^\pm . For Abelian gauge theory, we introduce a complex line bundle L and a connection A_μ on L . The *Weyl spinor* $M(\overline{M})$ is a section of $S^+ \otimes L$ ($S^+ \otimes L^{-1}$), hence M satisfies the *positive chirality condition* $\gamma^5 M = M$. If X does not have Spin structure, we introduce $Spin^c$ structure and $Spin^c$ bundles $S^\pm \otimes L$, where L^2 is a line bundle. In this case, M should be interpreted as a section of $S^+ \otimes L$. Below, we assume Spin structure.

Recall that the 4D Abelian SW monopole equations are the following set of differential equations

$$F_{\mu\nu}^+ + \frac{i}{2} \overline{M} \sigma_{\mu\nu} M = 0, \quad i\gamma^\mu D_\mu M = 0, \quad (6.191)$$

where $F_{\mu\nu}^+$ is the self-dual part of the $U(1)$ curvature tensor

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \quad F_{\mu\nu}^+ = P_{\mu\nu\rho\sigma}^+ F^{\rho\sigma}, \quad (6.192)$$

while $P_{\mu\nu\rho\sigma}^+$ is the self-dual projector defined by

$$P_{\mu\nu\rho\sigma}^+ = \frac{1}{2} \left(\delta_{\mu\rho} \delta_{\nu\sigma} + \frac{\sqrt{g}}{2} \epsilon_{\mu\nu\rho\sigma} \right).$$

Note that the second term in the first equation of (6.191) is also self-dual. On the other hand, the second equation in (6.191) is a *twisted Dirac equation*, whose covariant derivative D_μ is given by

$$D_\mu = \partial_\mu + \omega_\mu - iA_\mu, \quad \text{where} \quad \omega_\mu = \frac{1}{4} \omega_\mu^{\alpha\beta} [\gamma_\alpha, \gamma_\beta]$$

is the *spin connection 1-form* on X .

In order to perform a reduction to 3D, let us first assume that X is a *product manifold* of the form $X = Y \times [0, 1]$, where Y is a 3D compact manifold which has Spin structure. We may identify $t \in [0, 1]$ as a ‘time’ variable, or, we assume t as the zero-th coordinate of X , whereas x_i ($i = 1, 2, 3$) are the coordinates on (space manifold) Y . Then the metric is given by

$$ds^2 = dt^2 + g_{ij}dx^i dx^j.$$

The dimensional reduction is proceeded by assumning that all fields are independent of t . (Below, we suppress the volume factor \sqrt{g} of Y for simplicity.)

First, let us consider the Dirac equation. After the dimensional reduction, the Dirac equation will be

$$\gamma^i D_i M - i\gamma^0 A_0 M = 0.$$

As for the first monopole equation, using (6.192) we find that

$$F_{i0} + \frac{1}{2}\epsilon_{i0jk}F^{jk} = -i\overline{M}\sigma_{i0}M, \quad F_{ij} + \epsilon_{ijk0}F^{k0} = -i\overline{M}\sigma_{ij}M. \quad (6.193)$$

Since the above two equations are dual each other, the first one, for instance, can be reduced to the second one by a contraction with the totally anti-symmetric tensor. Thus, it is sufficient to consider one of them. Here, we take the first equation in (6.193).

After the dimensional reduction, (6.193) will be

$$\partial_i A_0 - \frac{1}{2}\epsilon_{ijk}F^{jk} = -i\overline{M}\sigma_{i0}M, \quad (6.194)$$

where we have set $\epsilon_{ijk} \equiv \epsilon_{0ijk}$.

Therefore, the 3D version of the SW equations are given by

$$\partial_i b - \frac{1}{2}\epsilon_{ijk}F^{jk} + i\overline{M}\sigma_{i0}M = 0, \quad i(\gamma^i D_i - i\gamma^0 b)M = 0, \quad (b \equiv A_0). \quad (6.195)$$

It is now easy to establish the non-Abelian 3D monopole equations as

$$\partial_i b^a + f_{abc}A_i^b b^c - \frac{1}{2}\epsilon_{ijk}F^{ajk} + i\overline{M}\sigma_{i0}T^a M = 0, \quad i(\gamma^i D_i - i\gamma^0 b)M = 0,$$

where we have abbreviated $\overline{M}\sigma_{\mu\nu}T^a M \equiv \overline{M}^i \sigma_{\mu\nu}(T^a)_{ij}M^j$, subscripts of $(T^a)_{ij}$ run from 1 to $\dim \mathfrak{g}$ and $b^a \equiv A_0^a$.

Next, let us find an action which produces (6.195). The simplest one is given by

$$S = \frac{1}{2} \int_Y \left[\left(\partial_i b - \frac{1}{2} \epsilon_{ijk} F^{jk} + i \overline{M} \sigma_{i0} M \right)^2 + |i(\gamma^i D_i - i\gamma^0 b)M|^2 \right] d^3x. \quad (6.196)$$

Note that the minimum of (6.196) is given by (6.195). In this sense, the 3D monopole equations are not equations of motion but rather of constraints. Furthermore, there is a constraint for b . To see this, let us rewrite (6.196) as

$$S = \int_Y d^3x \left[\frac{1}{2} \left(\frac{1}{2} \epsilon_{ijk} F^{jk} - i \overline{M} \sigma_{i0} M \right)^2 + \frac{1}{2} |\gamma^i D_i M|^2 + \frac{1}{4} (\partial_i b)^2 + \frac{1}{2} b^2 |M|^2 \right].$$

The minimum of this action is clearly given by the 3D monopole equations with $b = 0$, for non-trivial A_i and M . However, for trivial A_i and M , we may relax the condition $b = 0$ to $\partial_i b = 0$, i.e., b is in general a non-zero constant. This can be also seen from (6.194). Therefore, we get

$$\begin{aligned} \frac{1}{2} \epsilon_{ijk} F^{jk} - i \overline{M} \sigma_{i0} M &= 0, & i\gamma^i D_i M &= 0, \\ \text{with } b = 0 & \quad \text{or} \quad \partial_i b = 0, \end{aligned} \quad (6.197)$$

as an equivalent expression to (6.195), but we will rather use (6.195) for convenience. The Gaussian action will be used in the next subsection in order to construct a TQFT by Batalin–Vilkovisky quantization algorithm. The non-Abelian version of (6.196) and (6.197) would be obvious.

6.5.3.2 TQFTs of 3D Monopoles

In this subsection, we construct TQFTs associated with both the Abelian and non-Abelian 3D monopoles by Batalin–Vilkovisky quantization algorithm.

Abelian Case

A 3D action for the Abelian 3D monopoles was found by the direct dimensional reduction of the 4D one [Zhang *et al.* (1995)], but here we rather show that the 3D topological action can be directly constructed from the 3D monopole equations [Ohta (1998)].

Topological Bogomol'nyi Action

A *topological Bogomol'nyi action* was constructed by using *Batalin–Vilkovisky quantization* algorithm [Birmingham *et. al.* (1989)], or quantization of a magnetic charge [Baulieu and Grossman (1988)]. The former is based on the quantization of a certain Langevin equation ('Bogomol'nyi monopole equation') and the classical action is quadratic, but the latter is based on the 'quantization' of the pure topological invariant by using the Bogomol'nyi monopole equation as a gauge fixing condition.

In order to compare the action to be constructed with those of Bogomol'nyi monopoles [Birmingham *et. al.* (1989); Baulieu and Grossman (1988)], we take Batalin–Vilkovisky procedure (see also [Birmingham *et. al.* (1991)]).

In order to get the topological action associated with 3D monopoles, we introduce random Gaussian fields G_i and $\nu(\bar{\nu})$ and then start with the action

$$S_c = \frac{1}{2} \int_Y [(G_i - \partial_i b + \frac{1}{2} \epsilon_{ijk} F^{jk} - i\bar{M}\sigma_{i0}M)^2 + |(\nu - i\gamma^i D_i M - \gamma^0 b M)|^2] d^3x. \quad (6.198)$$

Note that G_i and $\nu(\bar{\nu})$ are also regarded as auxiliary fields. This action reduces to (6.196) in the gauge

$$G_i = 0, \quad \nu = 0. \quad (6.199)$$

Firstly, note that (6.198) is invariant under the topological gauge transformation

$$\begin{aligned} \delta A_i &= \partial_i \theta + \epsilon_i, & \delta b &= \tau, & \delta M &= i\theta M + \varphi, \\ \delta G_i &= \partial_i \tau - \epsilon_{ijk} \partial^j \epsilon^k + i(\bar{\varphi} \sigma_{i0} M + \bar{M} \sigma_{i0} \varphi), \\ \delta \nu &= i\theta \nu + \gamma^i \epsilon_i M + i\gamma^i D_i \varphi + \gamma^0 b \varphi + \gamma^0 \tau M, \end{aligned} \quad (6.200)$$

where θ is the parameter of gauge transformation, ϵ_i and $\tau \equiv \epsilon_4$ are parameters which represent the topological shifts and φ the shift on the spinor space. The brackets for indices means anti-symmetrization, i.e., $A_{[i}B_{j]} = A_i B_j - A_j B_i$.

Here, let us classify the gauge algebra (6.200). This is necessary to use Batalin–Vilkovisky algorithm. Let us recall that the local symmetry for fields ϕ_i can be written generally in the form $\delta \phi_i = R_\alpha^i(\phi) \epsilon^\alpha$, where the indices mean the label of fields and ϵ^α is a some local parameter. When $\delta \phi_i = 0$ for non-zero ϵ^α , this symmetry is called *first-stage reducible*. In the

reducible theory, we can find zero-eigenvectors Z_a^α satisfying $R_\alpha^i Z_a^\alpha = 0$. Moreover, when the theory is *on-shell reducible*, we can find such eigenvectors by using equations of motion.

For the case at hand, under the identifications

$$\begin{aligned} \theta &= \Lambda, & \epsilon_i &= -\partial_i \Lambda, & \varphi &= -i\Lambda M, \\ \text{and } \tau &= 0, & \text{so that (6.200) will be} & & \end{aligned} \quad (6.201)$$

$$\begin{aligned} \delta A_i &= 0, & \delta b &= 0, & \delta M &= 0, & \delta G_i &= 0, \\ \delta \nu &= i\Lambda(\nu - i\gamma^i D_i M - \gamma^0 b M)|_{on-shell} = 0. \end{aligned} \quad (6.202)$$

Then for δA_i , for example, the R coefficients and the zero-eigenvectors are derived from

$$\begin{aligned} \delta A_i &= R_\theta^{A_i} Z_\Lambda^\theta + R_{\epsilon_j}^{A_i} Z_\Lambda^{\epsilon_j} = 0, & \text{that is} \\ R_\theta^{A_i} &= \partial_i, & R_{\epsilon_j}^{A_i} &= \delta_{ij}, & Z_\Lambda^\theta &= 1, & Z_\Lambda^{\epsilon_j} &= -\partial_j. \end{aligned}$$

Obviously, similar relations hold for other fields. The reader may think that the choice (6.201) is not suitable as a first stage reducible theory, but note that the zero-eigenvectors appear on every point where the gauge equivalence and the topological shift happen to coincide. In this three dimensional theory, $b(A_0)$ is invariant for the usual infinitesimal gauge transformation because of its ‘time’ independence, so (6.201) means that the existence of the points on spinor space where the topological shift trivializes indicates the first stage reducibility.

If we carry out *BRST quantization* via *Faddeev–Popov procedure* in this situation, the Faddeev–Popov determinant will have zero modes. Therefore in order to fix the gauge further we need a ghost for ghost. This reflects on the second generation gauge invariance (6.202) realized on-shell. However, since b is irrelevant to Λ , the ghost for τ will not couple to the second generation ghost. With this in mind, we use Batalin–Vilkovisky algorithm in order to make BRST quantization (for details, see [Ohta (1998)] and references therein).

Let us assign new ghosts carrying opposite statistics to the local parameters. The assortment is given by

$$\theta \longrightarrow c, \quad \epsilon_i \longrightarrow \psi_i, \quad \tau \longrightarrow \xi, \quad \varphi \longrightarrow N, \quad \Lambda \longrightarrow \phi. \quad (6.203)$$

Ghosts in (6.203) are first generations, in particular, c is Faddeev–Popov ghost, whereas ϕ is a second generation ghost. Their Grassmann parity and

ghost number (U number) are given by

$$\begin{array}{cccccc} c & \psi_i & \xi & N & \phi \\ 1^- & 1^- & 1^- & 1^- & 2^+, \end{array} \quad (6.204)$$

where the superscript of ghost number denotes the Grassmann parity. Note that the ghost number counts the degree of differential form on the moduli space \mathcal{M} of the solution to the 3D monopole equations. The minimal set Φ_{min} of fields consists of

$$\begin{array}{cccccc} A_i & b & M & G_i & \nu \\ 0^+ & 0^+ & 0^+ & 0^+ & 0^+, \end{array} \quad \text{and} \quad (6.204).$$

On the other hand, the set of anti-fields Φ_{min}^* carrying opposite statistics to Φ_{min} is given by

$$\begin{array}{cccccccccc} A_i^* & b^* & M^* & G_i^* & \nu^* & c^* & \psi_i^* & N^* & \phi^* \\ -1^- & -1^- & -1^- & -1^- & -1^- & -2^+ & -2^+ & -2^+ & -3^- \end{array}.$$

Next step is to find a solution to the master equation with Φ_{min} and Φ_{min}^* , given by

$$\frac{\partial_r S}{\partial \Phi^A} \frac{\partial_l S}{\partial \Phi_A^*} - \frac{\partial_r S}{\partial \Phi_A^*} \frac{\partial_l S}{\partial \Phi^A} = 0, \quad (6.205)$$

where $r(l)$ denotes right (left) derivative.

The general solution for the first stage reducible theory at hand can be expressed by

$$S = S_c + \Phi_i^* R_\alpha^i C_1^\alpha + C_{1\alpha}^* (Z_\beta^\alpha C_2^\beta + T_{\beta\gamma}^\alpha C_1^\gamma C_1^\beta) + C_{2\gamma}^* A_{\beta\alpha}^\gamma C_1^\alpha C_2^\beta + \Phi_i^* \Phi_j^* B_\alpha^{ji} C_2^\alpha + \dots, \quad (6.206)$$

where $C_1^\alpha (C_2^\alpha)$ denotes generally the first (second) generation ghost and only relevant terms in our case are shown. We often use $\Phi_{min}^A = (\phi^i, C_1^\alpha, C_2^\beta)$, where ϕ^i denote generally the fields. In this expression, the indices should be interpreted as the label of fields. Do not confuse with space-time indices. The coefficients $Z_\beta^\alpha, T_{\beta\gamma}^\alpha$, etc can be directly determined from the master equation. In fact, it is known that these coefficients satisfy the following relations

$$\begin{aligned} R_\alpha^i Z_\beta^\alpha C_2^\beta - 2 \frac{\partial_r S_c}{\partial \phi^j} B_\alpha^{ji} C_2^\alpha (-1)^{|i|} &= 0, \quad \frac{\partial_r R_\alpha^i C_1^\alpha}{\partial \phi^j} R_\beta^j C_1^\beta + R_\alpha^i T_{\beta\gamma}^\alpha C_1^\gamma C_1^\beta = 0, \\ \frac{\partial_r Z_\beta^\alpha C_2^\beta}{\partial \phi^j} R_\gamma^j C_1^\gamma + 2 T_{\beta\gamma}^\alpha C_1^\gamma Z_\delta^\beta C_2^\delta + Z_\beta^\alpha A_{\delta\gamma}^\beta C_1^\gamma C_2^\delta &= 0, \end{aligned} \quad (6.207)$$

where $|i|$ means the Grassmann parity of the i th field.

In these expansion coefficients, R_α^i and Z_β^α are related to the local symmetry (6.200). On the other hand, as $T_{\beta\gamma}^\alpha$ is related to the structure constant of a given Lie algebra for a gauge theory, it is generally called as structure function. Of course if the theory is Abelian, such structure function does not appear. However, for a theory coupled with matters, all of the structure functions do not always vanish, even if the gauge group is Abelian. At first sight, this seems to be strange, but the expansion (6.206) obviously detects the coupling of matter fields and ghosts. In fact, the appearance of this type of structure function is required in order to make the action to be constructed being full BRST invariant.

After some algebraic works, we will find the solution to be

$$S(\Phi_{min}, \Phi_{min}^*) = S_c + \int_Y \Delta S d^3x, \quad \text{where}$$

$$\begin{aligned} \Delta S = & A_i^*(\partial^i c + \psi^i) + b^* \xi + M^*(icM + N) + \overline{M}^*(-ic\overline{M} + \overline{N}) \\ & + G_i^* [\partial^i \xi - \epsilon^{ijk} \partial_j \psi_k + i(\overline{N} \sigma^{i0} M + \overline{M} \sigma^{i0} N)] \\ & + \nu^*(ic\nu + i\gamma^i D_i N + \gamma^i \psi_i M + \gamma^0 bN + \gamma^0 \xi M) \\ & + \overline{\nu}^*(ic\nu + i\gamma^i D_i N + \gamma^i \psi_i M + \gamma^0 bN + \gamma^0 \xi M) \\ & + c^* \phi - \psi_i^* \partial^i \phi - iN^*(\phi M + cN) + i\overline{N}^*(\phi \overline{M} + c\overline{N}) + 2i\nu^* \overline{\nu}^* \phi. \end{aligned}$$

We augment Φ_{min} by new fields $\chi_i, d_i, \mu(\overline{\mu}), \zeta(\overline{\zeta}), \lambda, \rho, \eta, e$ and the corresponding anti-fields. Their ghost number and Grassmann parity are given by

$$\begin{array}{cccccccc} \chi_i & d_i & \mu & \zeta & \lambda & \rho & \eta & e \\ -1^- & 0^+ & -1^- & 0^+ & -2^+ & -1^- & -1^- & 0^+ \end{array},$$

and

$$\begin{array}{cccc} \chi_i^* & \mu^* & \lambda^* & \rho^* \\ 0^+ & 0^+ & 1^- & 0^+ \end{array}.$$

Then we look for the solution

$$S' = S(\Phi_{min}, \Phi_{min}^*) + \int_Y (\chi^{*i} d_i + \overline{\mu}^* \overline{\zeta} + \mu^* \zeta + \rho^* e + \lambda^* \eta) d^3x, \quad (6.208)$$

where d_i, ζ, e, η , are Lagrange multiplier fields.

In order to get the quantum action we must fix the gauge. The best choice for the gauge fixing condition, which can reproduce the action obtained from the dimensional reduction of the 4D one, is found to be [Ohta

(1998)]

$$G_i = 0, \quad \nu = 0, \quad \partial^i A_i = 0, \\ -\partial^i \psi_i + \frac{i}{2}(\overline{N}M - \overline{M}N) = 0.$$

Thus we can get the gauge fermion carrying the ghost number -1 and odd Grassmann parity,

$$\Psi = -\chi^i G_i - \overline{\mu}\nu - \mu\overline{\nu} + \rho\partial^i A_i - \lambda \left[-\partial^i \psi_i + \frac{i}{2}(\overline{N}M - \overline{M}N) \right].$$

The quantum action S_q can be obtained by eliminating anti-fields and are restricted to lie on the gauge surface $\Phi^* = \frac{\partial_r \Psi}{\partial \Phi}$. Therefore the anti-fields will be

$$\begin{aligned} G_i^* &= -\chi_i, & \chi_i^* &= -G_i, & \nu^* &= -\overline{\mu}, & \overline{\nu}^* &= -\mu, & \overline{\mu}^* &= -\nu, \\ \mu^* &= -\overline{\nu}, & M^* &= -\frac{i}{2}\lambda\overline{N}, & \overline{M}^* &= \frac{i}{2}\lambda N, & N^* &= \frac{i}{2}\lambda\overline{M}, \\ \overline{N}^* &= -\frac{i}{2}\lambda M, & \rho^* &= \partial^i A_i, & A_i^* &= -\partial_i \rho, & \psi_i^* &= -\partial_i \lambda, \\ \lambda^* &= -\left[-\partial^i \psi_i + \frac{i}{2}(\overline{N}M - \overline{M}N) \right], & c^* &= \phi^* = b^* = \zeta^*(\overline{\zeta}^*) = 0. \end{aligned} \quad (6.209)$$

Then the quantum action S_q is given by $S_q = S'(\Phi, \Phi^* = \partial_r \Psi / \partial \Phi)$. Substituting (6.209) into S_q , we find that

$$S_q = S_c + \int_Y \tilde{\Delta} S d^3 x, \quad \text{where}$$

$$\begin{aligned} \tilde{\Delta} S &= (-\Delta\phi + \phi\overline{M}M - i\overline{N}N)\lambda - \left[-\partial^i \psi_i + \frac{i}{2}(\overline{N}M - \overline{M}N) \right] \eta \\ &\quad - \overline{\mu}(i c \nu + i \gamma^i D_i N + \gamma^i \psi_i M + \gamma^0 b N + \gamma^0 \xi M) \\ &\quad + \overline{(i c \nu + i \gamma^i D_i N + \gamma^i \psi_i M + \gamma^0 b N + \gamma^0 \xi M)} \mu + 2i\phi\overline{\mu}\mu \\ &\quad - \chi^i \left[\partial_i \xi - \epsilon_{ijk} \partial^j \psi^k + i(\overline{N}\sigma_{i0}M + \overline{M}\sigma_{i0}N) \right] \\ &\quad + \rho(\Delta c + \partial^i \psi_i) - d^i G_i - \overline{\zeta}\nu - \overline{\nu}\zeta + e\partial^i A_i. \end{aligned} \quad (6.210)$$

Using the condition (6.199) with $c = 0$, we can arrive at

$$S'_q = S_c|_{G_i=\nu(\overline{\nu})=0} + \int_Y \tilde{\Delta} S|_{c=0} d^3 x, \quad \text{where} \quad (6.211)$$

$$\begin{aligned}
\tilde{\Delta}S|_{c=0} = & (-\Delta\phi + \phi\overline{M}M - i\overline{N}N)\lambda - \left[-\partial^i\psi_i + \frac{i}{2}(\overline{N}M - \overline{M}N) \right] \eta \\
& - \overline{\mu}(i\gamma^i D_i N + \gamma^i \psi_i M + \gamma^0 bN + \gamma^0 \xi M) \\
& + \overline{(i\gamma^i D_i N + \gamma^i \psi_i M + \gamma^0 bN + \gamma^0 \xi M)} \mu + 2i\phi\overline{\mu}\mu \\
& - \chi^i \left[\partial_i \xi - \epsilon_{ijk} \partial^j \psi^k + i(\overline{N}\sigma_{i0}M + \overline{M}\sigma_{i0}N) \right] + \rho \partial^i \psi_i + e \partial^i A_i.
\end{aligned}$$

It is easy to find that (6.211) is consistent with the action found by the dimensional reduction of the 4D topological action [Zhang *et. al.* (1995)].

BRST Transformation

The Batalin–Vilkovisky algorithm also facilitates to construct BRST transformation rule. The BRST transformation rule for a field Φ is defined by

$$\delta_B \Phi = \epsilon \left. \frac{\partial_r S'}{\partial \Phi^*} \right|_{\Phi^* = \frac{\partial_r \Psi}{\partial \Phi}}, \quad (6.212)$$

where ϵ is a constant Grassmann odd parameter. With this definition for (6.210), we get

$$\begin{aligned}
\delta_B A_i &= -\epsilon(\partial_i c + \psi_i), & \delta_B b &= -\epsilon\xi, & \delta_B M &= -\epsilon(icM + N), \\
\delta_B G_i &= -\epsilon \left[\partial_i \xi - \epsilon_{ijk} \partial^j \psi^k + i(\overline{N}\sigma_{i0}M + \overline{M}\sigma_{i0}N) \right], \\
\delta_B \nu &= -\epsilon(ic\nu + i\gamma^i D_i N + \gamma^i \psi_i M + \gamma^0 bN + \gamma^0 \xi M - i\mu\phi), \\
\delta_B C &= \epsilon\phi, & \delta_B \psi_i &= -\epsilon\partial_i \phi, & \delta_B \rho &= \epsilon e, & \delta_B \lambda &= -\epsilon\eta, \\
\delta_B \mu &= \epsilon\zeta, & \delta_B N &= -i\epsilon(\phi M + cN), & \delta_B \chi_i &= \epsilon d_i, \\
\delta_B \phi &= \delta_B \xi = \delta_B d_i = \delta_B e = \delta_B \zeta = \delta_B \eta = 0.
\end{aligned} \quad (6.213)$$

It is clear at this stage that (6.213) has on-shell nilpotency, i.e., the quantum equation of motion for ν must be used in order to have $\delta_B^2 = 0$. This is due to the fact that the gauge algebra has on-shell reducibility. Accordingly, the Batalin–Vilkovisky algorithm gives a BRST invariant action and on-shell nilpotent BRST transformation. Note that the equations

$$\begin{aligned}
\partial_i \xi - \epsilon_{ijk} \partial^j \psi^k + i(\overline{N}\sigma_{i0}M + \overline{M}\sigma_{i0}N) &= 0, \\
i\gamma^i D_i N + \gamma^i \psi_i M + \gamma^0 bN + \gamma^0 \xi M &= 0
\end{aligned}$$

can be recognized as linearizations of the 3D monopole equations and the number of linearly independent solutions gives the dimension of the moduli space \mathcal{M} .

It is now easy to show that the global supersymmetry can be recovered from (6.213). In Witten type theory, Q_B can be interpreted as a supersymmetric BRST charge. We define the supersymmetry transformation as [Ohta (1998)]

$$\delta_S \Phi := \delta_B \Phi|_{c=0}.$$

Off-Shell Action

As was mentioned before, the quantum action of Witten type TQFT can be represented by BRST commutator with nilpotent BRST charge Q_B . However, since our BRST transformation rule is on-shell nilpotent, we should integrate out ν and G_i in order to get off-shell BRST transformation and off-shell quantum action.

For this purpose, let us consider the following terms in (6.210),

$$\frac{1}{2}(G_i - X_i)^2 + \frac{1}{2}|\nu - A|^2 - i\bar{\mu}c\nu + i\bar{c}\nu\mu - \bar{\zeta}\nu - \bar{\nu}\zeta - d^i G_i, \quad (6.214)$$

$$\text{where} \quad X_i = \partial_i b - \frac{1}{2}\epsilon_{ijk}F^{jk} + i\bar{M}\sigma_{i0}M, \quad A = i\gamma^i D_i M + \gamma^0 bM.$$

Here, let us define

$$\nu' = \nu - A, \quad B = -i\bar{c}\mu - \zeta.$$

$\nu'(\bar{\nu}')$ and G_i can be integrated out and then (6.214) will be

$$-\frac{1}{2}d_i d^i - d_i X^i - 2|B|^2 + \bar{B}A + B\bar{A}.$$

Consequently, we get the off-shell quantum action

$$S_q = \{Q, \tilde{\Psi}\}, \quad \text{where} \quad (6.215)$$

$$\begin{aligned} \tilde{\Psi} = & -\chi^i \left(X_i + \frac{\alpha}{2} d_i \right) - \mu \overline{(i\gamma^i D_i M + \gamma^0 bM - \beta B)} \\ & - \bar{\mu} (i\gamma^i D_i M + \gamma^0 bM - \beta B) + \rho \partial^i A_i - \lambda \left[-\partial^i \psi_i + \frac{i}{2}(\bar{N}M - \bar{M}N) \right]. \end{aligned}$$

α and β are arbitrary gauge fixing paramaters. Convenience choice for them is $\alpha = \beta = 1$. The BRST transformation rule for X_i and B fields can be easily obtained, although we do not write down here.

Observables

We can now discuss the observables. For this purpose, let us define [Baulieu and Grossman (1988)]

$$\mathcal{A} = A + c, \quad \mathcal{F} = F + \psi - \phi, \quad \mathcal{K} = db + \xi,$$

where we have introduced differential form notations, but their meanings would be obvious. A and c are considered as a $(1, 0)$ and $(0, 1)$ part of 1-form on (Y, \mathcal{M}) . Similarly, F, ψ and ϕ are $(2, 0), (1, 1)$ and $(0, 2)$ part of the 2-form \mathcal{F} , and db and ξ are $(1, 0)$ and $(0, 1)$ part of the 1-form \mathcal{K} . Thus \mathcal{A} defines a connection 1-form on (Y, \mathcal{M}) and \mathcal{F} is a curvature 2-form. Note that the exterior derivative d maps any (p_1, p_2) -form X_p of total degree $p = p_1 + p_2$ to $(p_1 + 1, p_2)$ -form, but δ_B maps any (p_1, p_2) -form to $(p_1, p_2 + 1)$ -form. Also note that $X_p X_q = (-1)^{pq} X_q X_p$. Then the action of δ_B is

$$(d + \delta_B)\mathcal{A} = \mathcal{F}, \quad (d + \delta_B)b = \mathcal{K}. \quad (6.216)$$

\mathcal{F} and \mathcal{K} also satisfy the following Bianchi identities in Abelian theory:

$$(d + \delta_B)\mathcal{F} = 0, \quad (d + \delta_B)\mathcal{K} = 0. \quad (6.217)$$

Equations (6.216) and (6.217) mean anti-commuting property between the BRST variation δ_B and the exterior differential d , i.e., $\{\delta_B, d\} = 0$.

The BRST transformation rule in geometric sector can be easily read from (6.213), i.e., $\delta_B A, \delta_B \psi, \delta_B c$ and $\delta_B \phi$. (6.217) implies

$$(d + \delta_B)\mathcal{F}^n = 0, \quad (6.218)$$

and expanding the above expression by ghost number and form degree, we get the following $(i, 2n - i)$ -form $W_{n,i}$,

$$\begin{aligned} W_{n,0} &= \frac{\phi^n}{n!}, & W_{n,1} &= \frac{\phi^{n-1}}{(n-1)!} \psi, \\ W_{n,2} &= \frac{\phi^{n-2}}{2(n-2)!} \psi \wedge \psi - \frac{\phi^{n-1}}{(n-1)!} F, \\ W_{n,3} &= \frac{\phi^{n-3}}{6(n-3)!} \psi \wedge \psi \wedge \psi - \frac{\phi^{n-2}}{(n-2)!} F \wedge \psi, \end{aligned} \quad (6.219)$$

$$\begin{aligned} \text{where} \quad 0 &= \delta_B W_{n,0}, & dW_{n,0} &= \delta_B W_{n,1}, \\ dW_{n,1} &= \delta_B W_{n,2}, & dW_{n,2} &= \delta_B W_{n,3}, & dW_{n,3} &= 0. \end{aligned} \quad (6.220)$$

Picking a certain k -cycle γ as a representative and defining the integral

$$W_{n,k}(\gamma) = \int_{\gamma} W_{n,k},$$

we can easily prove that

$$\delta_B W_{n,k}(\gamma) = - \int_{\gamma} dW_{n,k-1} = - \int_{\partial\gamma} W_{n,k-1} = 0,$$

as a consequence of (6.220). Note that the last equality follows from the fact that the cycle γ is a simplex without boundary, i.e., $\partial\gamma = 0$. Therefore, $W_{n,k}(\gamma)$ indeed gives a topological invariant associated with n -th Chern class on $Y \times \mathcal{M}$.

On the other hand, since we have a scalar field b and its ghosts, we may construct topological observables associated with them. Therefore, the observables can be obtained from the ghost expansion of

$$(d + \delta_B) \mathcal{F}^n \wedge \mathcal{K}^m = 0.$$

Explicitly, for $m = 1$, for example, we get

$$\begin{aligned} 0 &= \delta_B W_{n,1,0}, & dW_{n,1,0} &= \delta_B W_{n,1,1}, & dW_{n,1,1} &= \delta_B W_{n,1,2}, \\ dW_{n,1,2} &= \delta_B W_{n,1,3}, & dW_{n,1,3} &= 0, & \text{where} \end{aligned}$$

$$\begin{aligned} W_{n,1,0} &= \frac{\phi^n}{n!} \xi, & W_{n,1,1} &= \frac{\phi^{n-1}}{(n-1)!} \psi \xi - \frac{\phi^n}{n!} db, \\ W_{n,1,2} &= \frac{\phi^{n-2}}{2(n-2)!} \psi \wedge \psi \xi - \frac{\phi^{n-1}}{(n-1)!} F \xi - \frac{\phi^{n-1}}{(n-1)!} \psi \wedge db, \\ W_{n,1,3} &= \frac{\phi^{n-3}}{6(n-3)!} \psi \wedge \psi \wedge \psi \xi + \frac{\phi^{n-1}}{(n-1)!} F \wedge db \\ &\quad + \frac{\phi^{n-2}}{2(n-2)!} (2\psi \wedge F \xi + \psi \wedge \psi \wedge db). \end{aligned} \quad (6.221)$$

These relations correspond to the cocycles [Baulieu and Grossman (1988)] in $U(1)$ case.

Next, let us look for the observables for matter sector. The BRST transformation rules in this sector is given by $\delta_B, \delta_B N, \delta_{BC}$ and $\delta_B \phi$. At first sight, the matter sector does not have any observable, but we can find the combined form

$$\widetilde{W} = i\phi \overline{M} M + \overline{N} N \quad (6.222)$$

is an observable. However, unfortunately, as \widetilde{W} is cohomologically trivial because $\delta_B \widetilde{W} = 0$ but $d\widetilde{W} \neq \delta_B \widetilde{W}'$ for some \widetilde{W}' . Accordingly, \widetilde{W} does not give any new topological invariant [Ohta (1998)].

In topological Bogomol'nyi theory, there is a sequence of observables associated with a magnetic charge. For the Abelian case, it is given by

$$W = \int_Y F \wedge db. \quad (6.223)$$

As is pointed out for the case of Bogomol'nyi monopoles [Birmingham *et. al.* (1989)], we can not get the observables related with this magnetic charge by the action of δ_B as well, but we can construct those observables by anti-BRST variation $\bar{\delta}_B$ which maps (m, n) -form to $(m, n-1)$ -form. $\bar{\delta}_B$ can be obtained by a discrete symmetry which is realized as 'time reversal symmetry' in 4D. In our 3D theory, the discrete symmetry is given by

$$\begin{aligned} \phi &\longrightarrow -\lambda, & \lambda &\longrightarrow -\phi, & N &\longrightarrow i\sqrt{2}\mu, & \mu &\longrightarrow \frac{i}{\sqrt{2}}N, \\ \psi_i &\longrightarrow \frac{\chi_i}{\sqrt{2}}, & \chi_i &\longrightarrow \sqrt{2}\psi_i, & \eta &\longrightarrow \sqrt{2}\xi, & \xi &\longrightarrow -\frac{\eta}{\sqrt{2}} \end{aligned} \quad (6.224)$$

$$\text{with } b \longrightarrow -b, \quad (6.225)$$

where (6.225) represents an additional symmetry [Birmingham *et. al.* (1989)]. Note that we must also change N and μ (and their conjugates). The positive chirality condition for M should be used in order to check the invariance of the action. In this way, we can get anti-BRST transformation rule by substituting (6.224) and (6.225) into (6.213) and then we can get the observables associated with the magnetic charge by using the action of this anti-BRST variation [Birmingham *et. al.* (1989)].

The topological observables available in this theory are the same with those of topological Bogomol'nyi monopoles.

Finally, let us briefly comment on our three dimensional theory. First note that Lagrangian L and Hamiltonian H in dimensional reduction can be considered as equivalent. This is because the relation between them is defined by

$$H = p\dot{q} - L,$$

where q is any field, the overdot means time derivative and p is a canonical conjugate momentum of q , and the dimensional reduction requires the time independence of all fields, thus $H = -L$ in this sense. Though we have constructed the three dimensional action directly from the 3D monopole

equations, our action may be interpreted essentially as the Hamiltonian of the four dimensional SW theory.

6.5.3.3 Non-Abelian Case

It is easy to extend the results obtained in the previous subsection to non-Abelian case. In this subsection, we summarize the results for the non-Abelian 3D monopoles (for details, see [Ohta (1998)] and references therein).

Non-Abelian Topological Action

With the auxiliary fields $G_{\mu\nu}^a$ and ν , we consider

$$S_c = \frac{1}{2} \int_Y d^3x [(G_i^a - K_i^a)^2 + |\nu - i\gamma^i D_i M - \gamma^0 b M|^2], \quad (6.226)$$

$$\text{where} \quad K_i^a = \partial_i b^a + f_{abc} A_i^b b^c - \frac{1}{2} \epsilon_{ijk} F_{jk}^a + i\overline{M} \sigma_{i0} T^a M.$$

Note that the minimum of (6.226) with the gauge $G_i^a = \nu = 0$ are given by the non-Abelian 3D monopoles. We take the generator of Lie algebra in the fundamental representation, e.g., for $SU(n)$,

$$(T_a)_{ij}(T^a)_{kl} = \delta_{il}\delta_{jk} - \frac{1}{n}\delta_{ij}\delta_{kl}.$$

Extension to other Lie algebra and representation is straightforward.

The gauge transformation rule for (6.226) is given by

$$\begin{aligned} \delta A_i^a &= \partial_i \theta^a + f_{abc} A_i^b \theta^c + \epsilon_i^a, & \delta b^a &= f_{abc} b^b \theta^c + \tau^a, & \delta M &= i\theta M + \varphi, \\ \delta G_i^a &= f_{abc} G_i^b \theta^c + [-\epsilon_{ijk}(\partial^j \epsilon^{ak} + f_{abc} \epsilon^{jb} A^{ck}) \\ &\quad + \partial_i \tau^a + f_{abc}(\epsilon_i^b b^c - \tau^b A_i^c) + i(\overline{\varphi} \sigma_{i0} T^a M + \overline{M} \sigma_{i0} T^a \varphi)], \\ \delta \nu &= i\gamma^i D_i \varphi + \gamma^i \epsilon_i M + \gamma^0 b \varphi + \gamma^0 \tau M + i\theta \nu. \end{aligned} \quad (6.227)$$

Note that we have a G_i^a term in the transformation of G_i^a , while it did not appear in Abelian theory.

The gauge algebra (6.227) possesses on-shell zero modes as in the Abelian case. Setting

$$\theta^a = \Lambda^a, \quad \epsilon_i^a = -\partial_i \Lambda^a - f_{abc} A_i^b \Lambda^c, \quad \tau^a = -f_{abc} b^b \Lambda^c, \quad \varphi = -i\Lambda M,$$

we can easily find that (6.227) closes, i.e.,

$$\begin{aligned}\delta A_i^a &= 0, & \delta b^a &= 0, & \delta M &= 0, \\ \delta G_i^a &= f_{abc}\Lambda^c[G_i^b - K_i^b]|_{on-shell} = 0, \\ \delta\nu &= i\Lambda[\nu - i(\gamma^i D_i - i\gamma^0 b)M]|_{on-shell} = 0,\end{aligned}\quad (6.228)$$

when the equations of motion of G_i^a and ν are used. Note that we must use both equations of motion of G_i^a and ν in the non-Abelian case, while only ‘ ν ’ was needed for the Abelian theory. Furthermore, as φ is a parameter in the spinor space, φ is not \mathfrak{g} -valued, in other words, $\varphi \neq \varphi^a T^a$. (6.227) is first stage reducible.

The assortment of ghost fields, the minimal set Φ_{min} of the fields and the ghost number and the Grassmann parity, furthermore those for Φ_{min}^* would be obvious.

Then the solution to the master equation will be

$$S(\Phi_{min}, \Phi_{min}^*) = S_c + \int_Y \text{Tr}(\Delta S_n) d^3x, \quad \text{where}$$

$$\begin{aligned}\Delta S_n &= A_i^*(D^i c + \psi^i) + b^*(i[b, c] + \xi) + M^*(icM + N) + \overline{M}^*(-ic\overline{M} + \overline{N}) + G_i^* \widetilde{G}^i \\ &- iN^*(\phi M + cN) + i\overline{N}^*(\phi\overline{M} + c\overline{N}) + \nu^*(ic\nu + i\gamma^i D_i N + \gamma^i \psi_i M + \gamma^0 bN + \gamma^0 \xi M) \\ &+ \overline{\nu}^*(ic\nu + i\gamma^i D_i N + \gamma^i \psi_i M + \gamma^0 bN + \gamma^0 \xi M) + 2i\nu^* \overline{\nu}^* \phi + \psi_i^*(-D^i \phi - i\{\psi^i, c\}) \\ &+ c^*\left(\phi - \frac{i}{2}\{c, c\}\right) - i\phi^*[\phi, c] - \frac{i}{2}\{G_i^*, G^{*i}\}\phi + i\xi^*([b, \phi] - \{\xi, c\}).\end{aligned}$$

Here

$$\widetilde{G}_i = i[c, G_i] - \epsilon_{ijk} D^j \psi^k + D_i \xi + [\psi_i, \xi] + i(\overline{N} \sigma_{i0} T_a T^a M + \overline{M} \sigma_{i0} T_a T^a N).$$

The equations

$$\begin{aligned}-\epsilon_{ijk} D^j \psi^k + D_i \xi + [\psi_i, \xi] + i(\overline{N} \sigma_{i0} T_a T^a M + \overline{M} \sigma_{i0} T_a T^a N) &= 0, \\ i\gamma^i D_i N + \gamma^i \psi_i M + \gamma^0 bN + \gamma^0 \xi M &= 0,\end{aligned}$$

can be seen as linearizations of non-Abelian 3D monopoles.

We augment Φ_{min} by new fields $\chi_i^a, d_i^a, \mu(\overline{\mu}), \zeta(\overline{\zeta}), \lambda, \rho, \eta, e$ and the corresponding anti-fields, but Lagrange multipliers fields $d_i^a, \zeta(\overline{\zeta}), e, \eta$, are assumed not to have anti-fields for simplicity and therefore their BRST transformation rules are set to zero. This simplification means that we do not take into account of BRST exact terms. In this sense, the result to be

obtained will correspond to those of the dimensionally reduced version of the 4D theory up to these terms, i.e., topological numbers.

From the gauge fixing condition

$$G_i^a = 0, \quad \nu = 0, \quad \partial^i A_i = 0, \quad -D^i \psi_i + \frac{i}{2}(\overline{N}M - \overline{M}N) = 0,$$

the gauge fermion will be

$$\Psi = -\chi^i G_i - \bar{\mu}\nu - \mu\bar{\nu} + \rho\partial^i A_i - \lambda \left[-D^i \psi_i + \frac{i}{2}(\overline{N}M - \overline{M}N) \right].$$

The anti-fields are then given by

$$\begin{aligned} G_i^* &= -\chi_i, \quad \chi_i^* = -G_i, \quad \nu^* = -\bar{\mu}, \quad \bar{\nu}^* = -\mu, \quad \bar{\mu}^* = -\nu, \quad \mu^* = -\bar{\nu}, \\ M^* &= -\frac{i}{2}\lambda\overline{N}, \quad \overline{M}^* = \frac{i}{2}\lambda N, \quad N^* = \frac{i}{2}\lambda\overline{M}, \quad \overline{N}^* = -\frac{i}{2}\lambda M, \\ \rho^* &= \partial^i A_i, \quad A_i^* = -\partial_i \rho + i[\lambda, \psi_i], \quad b^* = c^* = \xi^* = \phi^* = \zeta^*(\bar{\zeta}^*) = 0, \\ \lambda^* &= - \left[-D_i \psi^i + [b, \xi] + \frac{i}{2}(\overline{N}M - \overline{M}N) \right], \quad \psi_i^* = -D_i \lambda. \end{aligned}$$

Therefore we find the quantum action

$$S_q = S_c + \int_Y \text{Tr} \left(\tilde{\Delta} S_n \right) d^3 x, \quad \text{where} \quad (6.229)$$

$$\begin{aligned} \tilde{\Delta} S_n &= - \left[-D_i \psi^i + [b, \xi] + \frac{i}{2}(\overline{N}M - \overline{M}N) \right] \eta - \lambda(D_i D^i \phi + iD_i \{\psi^i, c\}), \\ &\quad + i\lambda\{\psi_i, D^i c + \psi^i\} + (\phi\overline{M}M - i\overline{N}N)\lambda, \\ -\chi^i &\left[i[c, G_i] + \epsilon_{ijk} D^j \psi^k + D_k \xi + [\psi_k, \xi] + \frac{i}{2}(\overline{N}\sigma^{ij} T_a T^a M + \overline{M}\sigma^{ij} T_a T^a N) \right], \\ &\quad -\bar{\mu}(i\gamma^\mu D_\mu N + \gamma^\mu \psi_\mu M + i\nu) + \overline{(i\gamma^\mu D_i N + \gamma^\mu \psi_\mu M + i\nu)}\mu, \\ &\quad + 2i\phi\bar{\mu}\mu - \frac{i}{2}\{\chi_i, \chi^i\}\phi + \rho(\partial_i D^i c + \partial_i \psi^i) - d^i G_i - \bar{\zeta}\nu - \bar{\nu}\zeta + e\partial^i A_i, \end{aligned} \quad (6.230)$$

In this quantum action, setting

$$M(\overline{M}) = N(\overline{N}) = \mu(\bar{\mu}) = \nu(\bar{\nu}) = 0,$$

we can find that the resulting action coincides with that of Bogomol'nyi monopoles [Birmingham *et. al.* (1989)].

Finally, in order to get the off-shell quantum action, both the auxiliary fields should be integrated out by the similar technique presented in Abelian case.

BRST transformation

The BRST transformation rule is given by

$$\begin{aligned}
 \delta_B A_i &= -\epsilon(D_i c + \psi_i), & \delta_B b &= -\epsilon(i[c, b] + \xi), & \delta_B \xi &= i\epsilon([b, \phi] - \{\xi, c\}), \\
 \delta_B M &= -\epsilon(icM + N), & \delta_B G_i &= -\epsilon(\tilde{G}_i - i[\chi_i, \phi]), \\
 \delta_B \nu &= -\epsilon(ic\nu + \gamma^\mu D_\mu N + \gamma^\mu \psi_\mu M - i\mu\phi), & \delta_B c &= \epsilon\left(\phi - \frac{i}{2}\{c, c\}\right), \\
 \delta_B \psi_i &= -\epsilon(D_i \phi + i\{\psi_i, c\}), & \delta_B \rho &= \epsilon e, & \delta_B \lambda &= -\epsilon\eta, \\
 \delta_B \mu &= \epsilon\zeta, & \delta_B N &= -i\epsilon(\phi M + cN), & \delta_B \chi_i &= \epsilon d_i, \\
 \delta_B \phi &= i\epsilon[\phi, c], & \delta_B d_i &= \delta_B e = \delta_B \zeta = \delta_B \eta = 0.
 \end{aligned} \tag{6.231}$$

It is easy to get supersymmetry also in this case. However, as we have omitted the BRST exact terms, the supersymmetry in our construction does not detect them.

Observables

We have already constructed the topological observables for Abelian case. Also in non-Abelian case, the construction of observables is basically the same. But the relation (6.216) and (6.217) are required to modify

$$(d + \delta_B)\mathcal{A} - \frac{i}{2}[\mathcal{A}, \mathcal{A}] = \mathcal{F}, \quad (d + \delta_B)b - i[\mathcal{A}, b] = \mathcal{K}, \quad \text{ar(6.232)}$$

$$(d + \delta_B)\mathcal{F} - i[\mathcal{A}, \mathcal{F}] = 0, \quad (d + \delta_B)\mathcal{K} - i[\mathcal{A}, \mathcal{K}] = i[\mathcal{F}, b], \quad (6.233)$$

respectively, where $[*, *]$ is a graded commutator. The observables in geometric and matter sector are the same as before, but we should replace db by $d_A b$ in (6.221) as well as (6.232) and (6.233), where d_A is a exterior covariant derivative and trace is required. In addition, the magnetic charge observables are again obtained by anti-BRST variation as outlined before.

The observables in geometric sector are those in (6.219) and follow the cohomological relation (6.220). In this way, the topological observables available in this three dimensional theory are precisely the Bogomol'nyi monopole cocycles [Baulieu and Grossman (1988)].

6.5.4 Stringy Actions and Amplitudes

Now we give a brief review of modern path-integral methods in superstring theory (mainly following [Deligne *et. al.* (1999)]). Recall that the funda-

mental quantities in quantum field theory (QFT) are the transition amplitudes $Amp : IN \Rightarrow OUT$, describing processes in which a number IN of incoming particles scatter to produce a number OUT of outgoing particles. The square modulus of the transition amplitude yields the *probability* for this process to take place.

6.5.4.1 Strings

Recall that in string theory, *elementary particles* are not described as 0-dimensional *points*, but instead as 1D *strings*. If M_s and $M(\sim \mathbb{R} \times M_s)$ denote the 3D space and 4D space-time manifolds respectively, then we picture strings as in Figure 6.14.

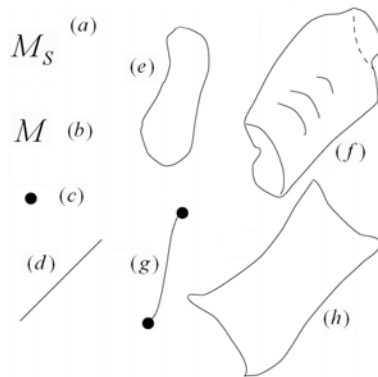


Fig. 6.14 Basic geometrical objects of string theory: (a) a space with fixed time; (b) a space-time picture; (c) a point-particle; (d) a world-line of a point-particle; (e) a closed string; (f) a world-sheet of a closed string; (g) an open string; (h) a world-sheet of an open string.

While the point-particle sweeps out a 1D *world-line*, the string sweeps out a *world-sheet*, i.e., a 2D real surface. For a *free string*, the topology of the world-sheet is a *cylinder* (in the case of a closed string) or a *sheet* (for an open string).

Roughly, different elementary particles correspond to different *vibration modes* of the string just as different minimal notes correspond to different vibrational modes of musical string instruments.

It turns out that the physical size of strings is set by *gravity*, more precisely the *Planck length* $\ell_P \sim 10^{-33}$ cm. This scale is so small that we effectively only see point-particles at our distance scales. Thus, for length scales much larger than ℓ_P , we expect to recover a QFT-description of

point-particles, plus typical *string corrections* that represent physics at the Planck scale.

6.5.4.2 Interactions

While the string itself is an extended 1D object, the fundamental string interactions are *local*, just as for point-particles. The interaction takes place when strings overlap in space at the same time. In case of *closed string theories* the interactions have a form depicted in Figure 6.15, while in case of *open string theories* the interactions have a form depicted in Figure 6.16. Other interactions result from combining the interactions defined above.

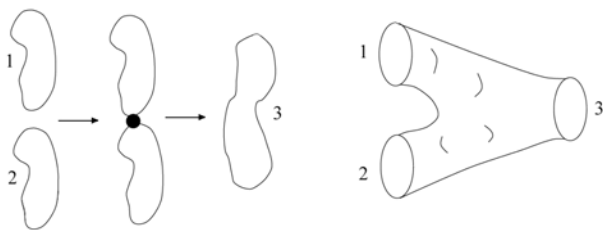


Fig. 6.15 Interactions in closed string theories (left 2D-picture and right 3D-picture).

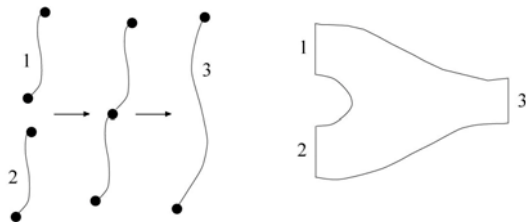


Fig. 6.16 Interactions in open string theories (left 2D-picture and right 3D-picture).

In point-particle theories, the fundamental interactions are read off from the QFT-Lagrangian. An interaction occurs at a *geometrical point*, where the world-lines join and cease to be a manifold. In *Lorentz-invariant theories* (where manifold M is a flat Minkowski space-time), the interaction point is Lorentz-invariant. To specify how the point-particles interact, additional data must be supplied at the interaction point, giving rise to *many possible* distinct quantum field theories.

In string theory, the interaction point *depends* upon the Lorentz frame

chosen to observe the process. In the Figure above, equal time slices are indicated from the point of view of two different Lorentz frames, schematically indicated by t and t' . The closed string interaction, as seen from frames t and t' , occurs at times t_2 and t'_2 and at (distinct) points P and P' respectively.

Lorentz invariance of interaction forbids that any point on the world-sheet be singled out as interaction point. Instead, the interaction results purely from the joining and splitting of strings. While free closed strings are characterized by their topology being that of a cylinder, interacting strings are characterized by the fact that their associated world-sheet is connected to at least 3 strings, incoming and/or outgoing.

As a result, the free string determines the nature of the interactions completely, leaving only the string coupling constant undetermined.

The *orientation* is an additional structure of closed strings, dividing them into two categories: (i) *oriented strings*, in which all world-sheets are assumed to be orientable; and (ii) non-oriented strings, in which world-sheets are non-orientable, such as the Möbius strip, Klein bottle, etc.

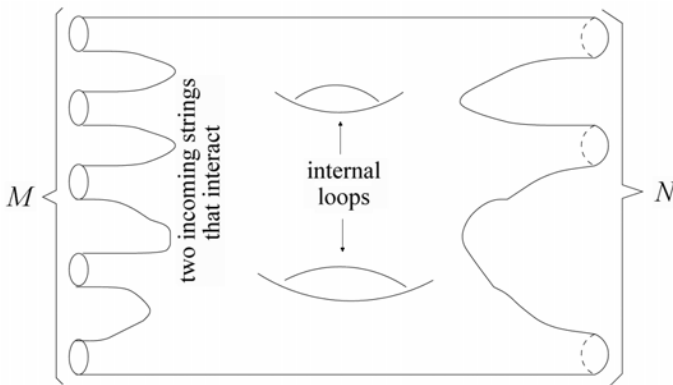


Fig. 6.17 Boundary components and handles of closed oriented system of M incoming strings, interacting through internal loops, to produce N outgoing strings. Note the striking similarity with MIMO-systems of *nonlinear control theory*, with M input processes and N output processes (see section 4.9.1 below).

6.5.4.3 Loop Expansion – Topology of Closed Surfaces

For simplicity, here we consider closed oriented strings only, so that the associated world-sheet is also oriented. A general string configuration de-

scribing the process in which M incoming strings interact and produce N outgoing strings looks at the topological level like a closed surface with $M + N = E$ *boundary components* and any number of *handles* (see Figure 6.17). This picture is a kind of topological generalization of nonlinear control MIMO-systems with M inputs, N outputs X states (see section 4.9.1 below).

The internal loops may arise when virtual particle pairs are produced, just as in quantum field theory. For example, a *Feynman diagram* in quantum field theory that involves a loop is shown in Figure 6.18 together with the corresponding string diagram.



Fig. 6.18 A QFT Feynman diagram that involves an internal loop (left), with the corresponding string diagram (right).

Surfaces associated with closed oriented strings have two topological invariants: (i) the *number of boundary components* $E = M + N$ (which may be shrunk to punctures, under certain conditions), and (ii) the number h of *handles* on the surface, which equals the surface *genus*.

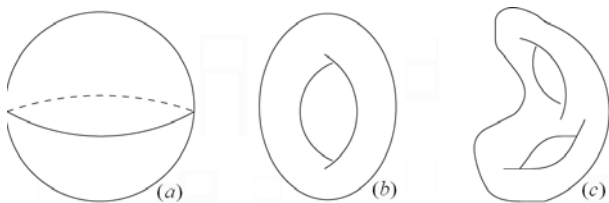


Fig. 6.19 Number h of handles on the surface of closed oriented strings, which equals the string-surface genus: (a) $h = 0$ for sphere S^2 ; (b) $h = 1$ for torus T^2 ; (c) $h = 2$ for string-surfaces with higher genus, etc.

When $E = 0$, we just have the topological classification of compact oriented surfaces without boundary. Rendering $E > 0$ is achieved by removing E discs from the surface.

Recall that in QFT, an expansion in powers of Planck's constant \hbar yields an expansion in the number of loops of the associated Feynman diagram,

for a given number of external states:

$$\hbar^{E+h-1} = \begin{cases} \hbar & \text{for every propagator} \\ \hbar^{-1} & \text{for every vertex} \\ -1 & \text{for overall momentum conservation} \end{cases}$$

Thus, in string theory we expect that, for a given number of external strings E , the topological expansion genus by genus should correspond to a *loop* expansion as well.

Recall that in QFT, there are in general many Feynman diagrams that correspond to an amplitude with a given number of external particles and a given number of loops. For example, for $E = 4$ external particles and $h = 1$ loop in ϕ^3 theory are given in Figure 6.20, together with the same process in string theory (for closed oriented strings), where it is described by just a single diagram (right).

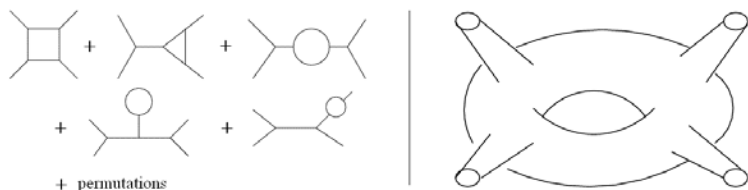


Fig. 6.20 Feynman QFT-diagrams for ϕ^3 theory with $E = 4$ external particles and $h = 1$ loop (left), and a single corresponding string diagram (right). In this way the usual Feynman diagrams of quantum field theory are generalized by arbitrary Riemannian surfaces.

Much of recent interest has been focused on the so-called D -branes. A D -brane is a submanifold of space-time with the property that strings can *end* or *begin* on it.

6.5.5 Transition Amplitudes for Strings

The only way we have today to define string theory is by giving a *rule* for the evaluation of transition amplitudes, order by order in the loop expansion, i.e., genus by genus. The rule is to assign a relative weight to a given configuration and then to sum over all configurations [Deligne *et al.* (1999)]. To make this more precise, we first describe the system's configuration manifold M (see Figure 6.21).



Fig. 6.21 The embedding map x from the reference surface Σ into the pseudo-Riemannian configuration manifold M (see text for explanation).

We assume that Σ and M are smooth manifolds, of dimensions 2 and n respectively, and that x is a continuous map from Σ to M . If ξ^m , (for $m = 1, 2$), are local coordinates on Σ and x^μ , ($\mu = 1, \dots, n$), are local coordinates on M then the map x may be described by functions $x^\mu(\xi^m)$ which are continuous.

To each system configuration we can associate a weight $e^{-S[x, \Sigma, M]}$, (for $S \in \mathbb{C}$) and the transition amplitude Amp for specified external strings (incoming and outgoing) is get by summing over all surfaces Σ and all possible maps x ,

$$Amp = \sum_{\text{surfaces } \Sigma} \sum_x e^{-S[x, \Sigma, M]}.$$

We now need to specify each of these ingredients:

(1) We assume M to be an nD Riemannian manifold, with metric g . A special case is flat Euclidean space-time \mathbb{R}^n . The space-time metric is assumed *fixed*.

$$ds^2 = (dx, dx)_g = g_{\mu\nu}(x) dx^\mu \otimes dx^\nu.$$

(2) The metric g on M induces a metric on Σ : $\gamma = x^*(g)$,

$$\gamma = \gamma_{mn} d\xi^m \otimes d\xi^n, \quad \gamma_{mn} = g_{\mu\nu} \frac{\partial x^\mu}{\partial \xi^m} \frac{\partial x^\nu}{\partial \xi^n}.$$

This metric is non-negative, but depends upon x . It is advantageous to introduce an intrinsic Riemannian metric g on Σ , independently of x ; in local coordinates, we have

$$g = g_{mn}(\xi) d\xi^m \otimes d\xi^n.$$

A natural intrinsic candidate for S is the area of $x(\Sigma)$, which gives the so-called *Nambu-Goto action*²⁰

$$\text{Area}(x(\Sigma)) = \int_{\Sigma} d\mu_{\gamma} = \int_{\Sigma} n^2 \xi \sqrt{\det \gamma_{mn}}, \quad (6.234)$$

which depends only upon g and x , but not on g [Goto (1971)]. However, the transition amplitudes derived from the Nambu-Goto action are *not* well-defined quantum-mechanically.

Otherwise, we can take as starting point the so-called *Polyakov action*²¹

$$S[x, g] = \kappa \int_{\Sigma} (dx, *dx)_g = \kappa \int_{\Sigma} d\mu_g g^{mn} \partial_m x^{\mu} \partial_n x^{\nu} g_{\mu\nu}(x), \quad (6.235)$$

where κ is the *string tension* (a positive constant with dimension of inverse length square). The stationary points of S with respect to g are at $g^0 = e^{\phi} \gamma$ for some function ϕ on Σ , and thus $S[x, g^0] \sim \text{Area}(x(\Sigma))$.

The Polyakov action leads to *well-defined* transition amplitudes, get by integration over the space $\text{Met}(\Sigma)$ of all positive metrics on Σ for a given topology, as well as over the space of all maps $\text{Map}(\Sigma, M)$. We can define the path integral

$$\text{Amp} = \sum_{\substack{\text{topologies} \\ \Sigma}} \int_{\text{Met}(\Sigma)} \frac{1}{N(g)} \int_{\text{Map}(\Sigma, M)} \mathcal{D}[x] e^{-S[x, g, g]},$$

where N is a normalization factor, while the measures $\mathcal{D}[g]$ and $\mathcal{D}[x]$ are constructed from $\text{Diff}^+(\Sigma)$ and $\text{Diff}(M)$ invariant L^2 norms on Σ and M . For fixed metric g , the action S is well-known: its stationary points are the harmonic maps $x: \Sigma \rightarrow M$ (see, e.g., [Eells and Lemaire (1978)]). However, g here varies and in fact is to be integrated over. For a general metric g , the action S defines a *nonlinear sigma model*, which is renormalizable because the dimension of Σ is 2. It would not in general be renormalizable in dimension higher than 2, which is usually regarded as an argument against the existence of fundamental membrane theories (see [Deligne *et al.* (1999)]).

²⁰Nambu-Goto action is the starting point of the analysis of string behavior, using the principles of ordinary Lagrangian mechanics. Just as the Lagrangian for a free point particle is proportional to its proper time i.e., the ‘length’ of its world-line, a relativistic string’s Lagrangian is proportional to the area of the sheet which the string traces as it travels through space-time.

²¹The Polyakov action is the 2D action from *conformal field theory*, used in string theory to describe the world-sheet of a moving string.

The Nambu–Goto action (6.234) and Polyakov action (6.235) represent the core of the so-called *bosonic string theory*, the original version of string theory, developed in the late 1960s. Although it has many attractive features, it also predicts a particle called the *tachyon* possessing some unsettling properties, and it has no fermions. All of its particles are *bosons*, the matter particles. The physicists have also calculated that bosonic string theory requires 26 space–time dimensions: 25 spatial dimensions and one dimension of time. In the early 1970s, *supersymmetry* was discovered in the context of string theory, and a new version of string theory called *superstring theory* (i.e., supersymmetric string theory) became the real focus, as it includes also *fermions*, the force particles. Nevertheless, bosonic string theory remains a very useful ‘toy model’ to understand many general features of *perturbative string theory* (see section 6.7 below).

6.5.6 Weyl Invariance and Vertex Operator Formulation

The action S is also invariant under *Weyl rescalings* of the metric g by a positive function on $\sigma: \Sigma \rightarrow \mathbb{R}$, given by $g \rightarrow e^{2\sigma}g$. In general, Weyl invariance of the full amplitude may be spoiled by anomalies. Assuming Weyl invariance of the full amplitude, the integral defining Amp may be *simplified* in two ways.

- 1) The integration over $Met(\Sigma)$ effectively collapses to an integration over the *moduli space of surfaces*, which is finite dimensional, for each genus h .
- 2) The boundary components of Σ — characterizing external string states — may be mapped to regular points on an underlying compact surface without boundary by conformal transformations. The *data*, such as momenta and other quantum numbers of the external states, are mapped into *vertex operators*. The amplitudes are now given by the path integral

$$Amp = \sum_{h=0}^{\infty} \int_{Met(\Sigma)} \mathcal{D}[g] \frac{1}{N(g)} \int_{Map(\Sigma, M)} \mathcal{D}[x] V_1 \dots V_N e^{-S},$$

for suitable vertex operators V_1, \dots, V_N .

6.5.7 More General Actions

Generalizations of the action S given above are possible when M carries extra structure.

1) M carries a 2-form $B \in \Omega^{(2)}(M)$. The resulting contribution to the action is also that of a ‘nonlinear sigma model’

$$S_B[x, B] = \int_{\Sigma} x^*(B) = \int_{\Sigma} dx^{\mu} \wedge dx^{\nu} B_{\mu\nu}(x)$$

2) M may carry a *dilaton field* $\Phi \in \Omega^{(0)}(M)$ so that

$$S_{\Phi}[x, \Phi] = \int_{\Sigma} d\mu_g R_g \Phi(x).$$

where R_g is the Gaussian curvature of Σ for the metric g .

3) There may be a *tachyon field* $T \in \Omega^{(0)}(M)$ contributing

$$S_T[x, T] = \int_{\Sigma} d\mu_g T(x).$$

6.5.8 Transition Amplitude for a Single Point Particle

The transition amplitude for a single point-particle could in fact be get in a way analogous to how we prescribed string amplitudes. Let space-time be again a Riemannian manifold M , with metric g . The prescription for the transition amplitude of a particle travelling from a point $y \in M$ to a point y' to M is expressible in terms of a sum over all (continuous) paths connecting y to y' :

$$Amp(y, y') = \sum_{\substack{\text{paths} \\ \text{joining } y \text{ and } y'}} e^{-S[path]}.$$

Paths may be parametrized by maps from $C = [0, 1]$ into M with $x(0) = y$, $x(1) = y'$. A simple world-line action for a massless particle is get by introducing a metric g on $[0, 1]$

$$S[x, g] = \frac{1}{2} \int_C d\tau g(\tau)^{-1} \dot{x}^{\mu} \dot{x}^{\nu} g_{\mu\nu}(x),$$

which is invariant under $Diff^+(C)$ and $Diff(M)$.

Recall that the analogous prescription for the point-particle transition amplitude is the path integral

$$Amp(y, y') = \int_{Met(C)} \mathcal{D}[g] \frac{1}{N} \int_{Map(C, M)} \mathcal{D}[x] e^{-S[x, g]}.$$

Note that for string theory, we had a prescription for transition amplitudes valid for all topologies of the world-sheet. For point-particles, there is only the topology of the interval C , and we can only describe a single point-particle, but not interactions with other point-particles. To put those in, we would have to supply additional information.

Finally, it is very instructive to work out the amplitude Amp by carrying out the integrations. The only $Diff^+(C)$ invariant of g is the length $L = \int_0^1 d\tau g(\tau)$; all else is generated by $Diff^+(C)$. Defining the normalization factor to be the volume of $Diff(C)$: $N = \text{Vol}(Diff(C))$ we have $\mathcal{D}[g] = \mathcal{D}[v] dL$ and the transition amplitude becomes

$$Amp(y, y') = \int_0^\infty dL \int \mathcal{D}[x] e^{-\frac{1}{2L} \int_0^1 d\tau (\dot{x}, \dot{x})_g} = \int_0^\infty dL \langle y' | e^{-L\Delta} | y \rangle = \left\langle y' \left| \frac{1}{\Delta} \right| y \right\rangle.$$

Thus, the amplitude is just the Green function at (y, y') for the Laplacian Δ and corresponds to the propagation of a massless particle (see [Deligne *et. al.* (1999)]).

6.5.9 Witten's Open String Field Theory

Noncommutative nature of space-time has often appeared in non-perturbative aspects of string theory. It has been used in a formulation of interacting open string field theory by Ed Witten [Witten (1986b); Witten (1986a)]. Witten has written a classical action of open string field theory in terms of noncommutative geometry, where the noncommutativity appears in a product of string fields. Later, the *Dirichlet branes* (or, D-branes) have been recognized as solitonic objects in superstring theory [Polchinski (1995)]. Further, it has been found that the low energy behavior of the D-branes are well described by supersymmetric Yang-Mills theory (SYM) [Witten (1996)]. In the situation of some D-branes coinciding, the space-time coordinates are promoted to matrices which appear as the fields in SYM. Then the size of the matrices corresponds to the number of the D-branes, so noncommutativity of the matrices is related to the noncommutative nature of space-time.

In this subsection, mainly following [Sugino (2000)], we review some basic properties of Witten's bosonic open string field theory [Witten (1986b)] and its explicit construction based on a *Fock space* representation of string field functional and δ -function overlap vertices [Gross and Jevicki(1987a); Gross and Jevicki(1987b); Cremmer *et. al.* (1986)].

Witten introduced a beautiful formulation of open string field theory in

terms of a noncommutative extension of differential geometry, where string fields, the BRST operator Q and the integration over the string configurations \int in string field theory are analogs of differential forms, the exterior derivative d and the integration over the manifold \int_M in the differential geometry, respectively. The ghost number assigned to the string field corresponds to the degree of the differential form. Also the (noncommutative) product between the string fields $*$ is interpreted as an analog of the wedge product \wedge between the differential forms.

The axioms obeyed by the system of \int , $*$ and Q are

$$\begin{aligned} \int QA &= 0, & Q(A * B) &= (QA) * B + (-1)^{n_A} A * (QB), \\ (A * B) * C &= A * (B * C), & \int A * B &= (-1)^{n_A n_B} \int B * A, \end{aligned}$$

where A , B and C are arbitrary string fields, whose ghost number is half-integer valued: The *ghost number* of A is defined by the integer n_A as $n_A + \frac{1}{2}$.

Then Witten discussed the following *string-field-theory action*

$$S = \frac{1}{G_s} \int \left(\frac{1}{2} \psi * Q\psi + \frac{1}{3} \psi * \psi * \psi \right), \quad (6.236)$$

where G_s is the open string coupling constant and ψ is the string field with the ghost number $-\frac{1}{2}$. The action is invariant under the gauge transformation

$$\delta\psi = Q\Lambda + \psi * \Lambda - \Lambda * \psi,$$

with the gauge parameter Λ of the ghost number $-\frac{3}{2}$.

6.5.9.1 Operator Formulation of String Field Theory

The objects defined above can be explicitly constructed by using the operator formulation, where the string field is represented as a Fock space, and the integration \int as an inner product on the Fock space. It was considered by [Gross and Jevicki(1987a); Gross and Jevicki(1987b)] in the case of the *Neumann boundary condition*. We will heavily use the notation of [Gross and Jevicki(1987a); Gross and Jevicki(1987b)]. In the operator formulation, the action (6.236) is described as

$$S = \frac{1}{G_s} \left(\frac{1}{2_{12}} \langle V_2 || \psi \rangle_1 Q | \psi \rangle_2 + \frac{1}{3_{123}} \langle V_3 || \psi \rangle_1 | \psi \rangle_2 | \psi \rangle_3 \right), \quad (6.237)$$

where the structure of the product $*$ in the kinetic and potential terms is encoded to that of the overlap vertices $\langle V_2|$ and $\langle V_3|$ respectively (here, subscripts put to vectors in the Fock space label the strings concerning the vertices).

As a preparation for giving the explicit form of the overlaps, let us consider open strings in 26-dimensional space-time with the constant metric G_{ij} in the Neumann boundary condition. The world sheet action is given by

$$S_{WS} = \frac{1}{4\pi\alpha'} \int d\tau \int_0^\pi d\sigma G_{ij} (\partial_\tau X^i \partial_\tau X^j - \partial_\sigma X^i \partial_\sigma X^j) + S_{gh}, \quad (6.238)$$

where S_{gh} is the action of the bc -ghosts:

$$S_{gh} = \frac{i}{2\pi} \int d\tau \int_0^\pi d\sigma [c_+ (\partial_\tau - \partial_\sigma) b_+ + c_- (\partial_\tau + \partial_\sigma) b_-]. \quad (6.239)$$

Under the Neumann boundary condition, the string coordinates have the standard mode expansions:

$$X^j(\tau, \sigma) = x^j + 2\alpha' \tau p^j + i\sqrt{2\alpha'} \sum_{n \neq 0} \frac{1}{n} \alpha_n^j e^{-in\tau} \cos(n\sigma), \quad (6.240)$$

also the mode expansions of the ghosts are given by

$$\begin{aligned} c_\pm(\tau, \sigma) &= \sum_{n \in \mathbb{Z}} c_n e^{-in(\tau \pm \sigma)} \equiv c(\tau, \sigma) \pm i\pi_b(\tau, \sigma), \\ b_\pm(\tau, \sigma) &= \sum_{n \in \mathbb{Z}} b_n e^{-in(\tau \pm \sigma)} \equiv \pi_c(\tau, \sigma) \mp ib(\tau, \sigma). \end{aligned}$$

As a result of the quantization, the modes obey the commutation relations:

$$[x^i, p^j] = iG^{ij}, \quad [\alpha_n^i, \alpha_m^j] = nG^{ij} \delta_{n+m, 0}, \quad \{b_n, c_m\} = \delta_{n+m, 0},$$

while the other terms vanish.

The overlap $|V_N\rangle = |V_N\rangle^X |V_N\rangle^{gh}$, ($N = 1, 2, \dots$) is the state satisfying the continuity conditions for the string coordinates and the ghosts at the N -string vertex of the string field theory. The superscripts X and gh show the contribution of the sectors of the coordinates and the ghosts respectively. The continuity conditions for the coordinates are

$$(X^{(r)j}(\sigma) - X^{(r-1)j}(\pi - \sigma)) |V_N\rangle^X = 0, \quad (P_i^{(r)}(\sigma) + P_i^{(r-1)}(\pi - \sigma)) |V_N\rangle^X = 0, \quad (6.241)$$

for $0 \leq \sigma \leq \frac{\pi}{2}$ and $r = 1, \dots, N$. Here $P_i(\sigma)$ is the momentum conjugate to the coordinate $X^j(\sigma)$ at $\tau = 0$, and the superscript (r) labels the string

(r) meeting at the vertex. In the above formulas, we regard $r = 0$ as $r = N$ because of the cyclic property of the vertex. For the ghost sector, we impose the following conditions on the variables $c(\sigma)$, $b(\sigma)$ and their conjugate momenta $\pi_c(\sigma)$, $\pi_b(\sigma)$:

$$\begin{aligned} (\pi_c^{(r)}(\sigma) - \pi_c^{(r-1)}(\pi - \sigma))|V_N\rangle^{gh} &= 0, & (b^{(r)}(\sigma) - b^{(r-1)}(\pi - \sigma))|V_N\rangle^{gh} &= 0, \\ (c^{(r)}(\sigma) + c^{(r-1)}(\pi - \sigma))|V_N\rangle^{gh} &= 0, & (\pi_b^{(r)}(\sigma) + \pi_b^{(r-1)}(\pi - \sigma))|V_N\rangle^{gh} &= 0, \end{aligned}$$

for $0 \leq \sigma \leq \frac{\pi}{2}$ and $r = 1, \dots, N$.

6.5.9.2 Open Strings in Constant B -Field Background

We consider a constant background of the second-rank antisymmetric tensor field B_{ij} in addition to the constant metric g_{ij} where open strings propagate. Then the boundary condition at the end points of the open strings changes from the Neumann type, and thus the open string has a different mode expansion from the Neumann case (6.240). As a result, the end point is to be noncommutative, in the picture of the D-branes which implies noncommutativity of the world volume coordinates on the D-branes. Here we derive the mode-expanded form of the open string coordinates as a preparation for a calculation of the overlap vertices in the next section.

We start with the world sheet action

$$\begin{aligned} S_{WS}^B &= \frac{1}{4\pi\alpha'} \int d\tau \int_0^\pi d\sigma [g_{ij}(\partial_\tau X^i \partial_\tau X^j - \partial_\sigma X^i \partial_\sigma X^j) \\ &\quad - 2\pi\alpha' B_{ij}(\partial_\tau X^i \partial_\sigma X^j - \partial_\sigma X^i \partial_\tau X^j)] + S_{gh}. \end{aligned} \quad (6.242)$$

Because the term proportional to B_{ij} can be written as a total derivative term, it does not affect the equation of motion but does the boundary condition, which requires

$$g_{ij} \partial_\sigma X^j - 2\pi\alpha' B_{ij} \partial_\tau X^j = 0 \quad (6.243)$$

on $\sigma = 0, \pi$. This can be rewritten to the convenient form

$$\begin{aligned} E_{ij} \partial_- X^j &= (E^T)_{ij} \partial_+ X^j, \\ \text{where } E_{ij} &\equiv g_{ij} + 2\pi\alpha' B_{ij}, \end{aligned} \quad (6.244)$$

and ∂_\pm are derivatives with respect to the light cone variables $\sigma^\pm = \tau \pm \sigma$. We can easily see that $X^j(\tau, \sigma)$ satisfying the boundary condition (6.244)

has the following mode expansion:

$$X^j(\tau, \sigma) = \tilde{x}^j + \alpha' [(E^{-1})^{jk} g_{kl} p^l \sigma^- + (E^{-1T})^{jk} g_{kl} p^l \sigma^+] \quad (6.245)$$

$$+ i\sqrt{\frac{\alpha'}{2}} \sum_{n \neq 0} \frac{1}{n} [(E^{-1})^{jk} g_{kl} \alpha_n^l e^{-in\sigma^-} + (E^{-1T})^{jk} g_{kl} \alpha_n^l e^{-in\sigma^+}].$$

We will get the commutators between the modes from the propagator of the open strings, which gives another derivation different from the method by [Chu and Ho (1999)] based on the quantization via the Dirac bracket. When performing the *Wick rotation*: $\tau \rightarrow -i\tau$ and mapping the world sheet to the upper half plane

$$z = e^{\tau+i\sigma}, \quad \bar{z} = e^{\tau-i\sigma} (0 \leq \sigma \leq \pi),$$

the boundary condition (6.244) becomes

$$E_{ij} \partial_{\bar{z}} X^j = (E^T)_{ij} \partial_z X^j, \quad (6.246)$$

which is imposed on the real axis $z = \bar{z}$. The propagator $\langle X^i(z, \bar{z}) X^j(z', \bar{z}') \rangle$ satisfying the boundary condition (6.246) is determined as

$$\langle X^i(z, \bar{z}) X^j(z', \bar{z}') \rangle = -\alpha' [g^{ij} \ln |z - z'| - g^{ij} \ln |z - \bar{z}'|$$

$$+ G^{ij} \ln |z - \bar{z}'|^2 + \frac{1}{2\pi\alpha'} \theta^{ij} \ln \frac{z - \bar{z}'}{\bar{z} - z'} + D^{ij},$$

where G^{ij} and θ^{ij} are given by

$$G^{ij} = \frac{1}{2} (E^{T-1} + E^{-1})^{ij} = (E^{T-1} g E^{-1})^{ij} = (E^{-1} g E^{T-1})^{ij}, \quad (6.247)$$

$$\theta^{ij} = 2\pi\alpha' \cdot \frac{1}{2} (E^{T-1} - E^{-1})^{ij} = (2\pi\alpha')^2 (E^{T-1} B E^{-1})^{ij} \quad (6.248)$$

$$= -(2\pi\alpha')^2 (E^{-1} B E^{T-1})^{ij}.$$

Also the constant D^{ij} remains unknown from the boundary condition alone. However it is an irrelevant parameter, so we can fix an appropriate value. The mode-expanded form (6.245) is mapped to

$$X^j(z, \bar{z}) = \tilde{x}^j - i\alpha' [(E^{-1})^{jk} p_k \ln \bar{z} + (E^{-1T})^{jk} p_k \ln z]$$

$$+ i\sqrt{\frac{\alpha'}{2}} \sum_{n \neq 0} \frac{1}{n} [(E^{-1})^{jk} \alpha_{n,k} \bar{z}^{-n} + (E^{-1T})^{jk} \alpha_{n,k} z^{-n}].$$

Note that the indices of p^l and α_n^l were lowered by the metric g_{ij} not G_{ij} . Recall the definition of the propagator

$$\langle X^i(z, \bar{z}) X^j(z', \bar{z}') \rangle \equiv R(X^i(z, \bar{z}) X^j(z', \bar{z}')) - N(X^i(z, \bar{z}) X^j(z', \bar{z}')), \quad (6.249)$$

where R and N stand for the radial ordering and the normal ordering respectively. We take a prescription for the normal ordering which pushes p_i to the right and \tilde{x}_j to the left with respect to the zero-modes p_i and \tilde{x}_j . It corresponds to considering the vacuum satisfying

$$p_j|0\rangle = \alpha_{n,j}|0\rangle = 0 \quad (n > 0), \quad \langle 0|\alpha_{n,j} = 0 \quad (n < 0), \quad (6.250)$$

which is the standard prescription for calculating the propagator of the massless scalar field in 2D conformal field theory from the operator formalism. Making use of (6.249), (6.250) and techniques of the contour integration, it is easy to get the commutators

$$[\alpha_{n,i}, \alpha_{m,j}] = n\delta_{n+m,0}G_{ij}, \quad [\tilde{x}^i, p_j] = i\delta_j^i,$$

where the first equation holds for all integers with $\alpha_{0,i} \equiv \sqrt{2\alpha'}p_i$. The constant D^{ij} is written as $\alpha'D^{ij} = -\langle 0|\tilde{x}^i\tilde{x}^j|0\rangle$. Let us fix D^{ij} as $\alpha'D^{ij} = -\frac{i}{2}\theta^{ij}$, which is the convention taken in [Seiberg and Witten (1999)]. Then the coordinates \tilde{x}^i become noncommutative:

$$[\tilde{x}^i, \tilde{x}^j] = i\theta^{ij},$$

but the center of mass coordinates $x^i \equiv \tilde{x}^i + \frac{1}{2}\theta^{ij}p_j$ can be seen to commute each other.

Now we have the mode-expanded form of the string coordinates and the commutation relations between the modes, which are

$$\begin{aligned} X^j(\tau, \sigma) &= x^j + 2\alpha' \left(G^{jk}\tau + \frac{1}{2\pi\alpha'}\theta^{jk}(\sigma - \frac{\pi}{2}) \right) p_k \\ &+ i\sqrt{2\alpha'} \sum_{n \neq 0} \frac{1}{n} e^{-in\tau} \left[G^{jk} \cos(n\sigma) - i\frac{1}{2\pi\alpha'}\theta^{jk} \sin(n\sigma) \right] \alpha_{n,k}, \end{aligned}$$

$$[\alpha_{n,i}, \alpha_{m,j}] = n\delta_{n+m,0}G_{ij}, \quad [x^i, p_j] = i\delta_j^i,$$

with all the other commutators vanishing.

Also, due to the formula

$$\sum_{n=1}^{\infty} \frac{2}{n} \sin(n(\sigma + \sigma')) = \begin{cases} \pi - \sigma - \sigma', & (\sigma + \sigma' \neq 0, 2\pi) \\ 0, & (\sigma + \sigma' = 0, 2\pi), \end{cases}$$

we can see by a direct calculation that the end points of the string become noncommutative

$$[X^i(\tau, \sigma), X^j(\tau, \sigma')] = \begin{cases} i\theta^{ij}, & (\sigma = \sigma' = 0) \\ -i\theta^{ij}, & (\sigma = \sigma' = \pi) \\ 0, & (\text{otherwise}). \end{cases}$$

On the other hand, it is noted that the conjugate momenta have the mode expansion identical with that in the Neumann case:

$$\begin{aligned} P_i(\tau, \sigma) &= \frac{1}{2\pi\alpha'} (g_{ij}\partial_\tau - 2\pi\alpha' B_{ij}\partial_\sigma) X^j(\tau, \sigma) \\ &= \frac{1}{\pi} p_i + \frac{1}{\pi\sqrt{2\alpha'}} \sum_{n \neq 0} e^{-in\tau} \cos(n\sigma) \alpha_{n,i}. \end{aligned}$$

Note that the relations (6.247) and (6.248) are in the same form as a T-duality transformation, although the correspondence is a formal sense, because we are not considering any compactification of space-time. The generalized T-duality transformation, namely $O(D, D)$ -transformation, is defined by

$$E' = (aE + b)(cE + d)^{-1}, \quad (6.251)$$

with a , b , c and d being $D \times D$ real matrices. (D is the dimension of space-time.) The matrix $h = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is $O(D, D)$ matrix, which satisfies

$$h^T J h = J, \quad \text{where} \quad J = \begin{pmatrix} 0 & \mathbf{1}_D \\ \mathbf{1}_D & 0 \end{pmatrix}.$$

The relations (6.247) and (6.248) correspond to the case of the inversion $a = d = 0$, $b = c = \mathbf{1}_D$.

6.5.9.3 Construction of Overlap Vertices

Here we construct Witten's open string theory in the constant B -field background by obtaining the explicit formulas of the overlap vertices. As is understood from the fact that the action of the ghosts (6.239) contains no background fields, the ghost sector is not affected by turning on the B -field background. Thus we may consider the coordinate sector only. First, let us

see the mode-expanded forms of the coordinates and the momenta at $\tau = 0$

$$\begin{aligned} X^j(\sigma) &= G^{jk} y_k + \frac{1}{\pi} \theta^{jk} \left(\sigma - \frac{\pi}{2} \right) p_k \\ &\quad + 2\sqrt{\alpha'} \sum_{n=1}^{\infty} \left[G^{jk} \cos(n\sigma) x_{n,k} + \frac{1}{2\pi\alpha'} \theta^{jk} \sin(n\sigma) \frac{1}{n} p_{n,k} \right], \\ P_i(\sigma) &= \frac{1}{\pi} p_i + \frac{1}{\pi\sqrt{\alpha'}} \sum_{n=1}^{\infty} \cos(n\sigma) p_{n,i}, \end{aligned}$$

where $x^j = G^{jk} y_k$, the coordinates and the momenta for the oscillator modes are

$$\begin{aligned} x_{n,k} &= \frac{i}{2} \sqrt{\frac{2}{n}} (a_{n,k} - a_{n,k}^\dagger) = \frac{i}{\sqrt{2n}} (\alpha_{n,k} - \alpha_{-n,k}), \\ p_{n,k} &= \sqrt{\frac{n}{2}} (a_{n,k} + a_{n,k}^\dagger) = \frac{1}{\sqrt{2}} (\alpha_{n,k} + \alpha_{-n,k}). \end{aligned}$$

The nonvanishing commutators are given by

$$[x_{n,k}, p_{m,l}] = iG_{kl} \delta_{n,m}, \quad [y_k, p_l] = iG_{kl}. \quad (6.252)$$

We should note that the metric appearing in eqs. (6.252) is G_{ij} , instead of g_{ij} . So it can be seen that if we employ the variables with the lowered space-time indices y_k , p_k , $x_{n,k}$ and $p_{n,k}$, the metric used in the expression of the overlaps is G^{ij} not g^{ij} .

The continuity condition (6.241) is universal for any background, and the mode expansion of the momenta $P_i(\sigma)$'s is of the same form as in the Neumann case, thus the continuity conditions for the momenta in terms of the modes $p_{n,i}$ are identical with those in the Neumann case. Also, since $p_{n,i}$'s mutually commute, it is natural to find a solution of the continuity condition, assuming the following form for the overlap vertices:

$$|\hat{V}_N\rangle_{1\dots N}^X = \exp \left[\frac{i}{4\pi\alpha} \theta^{ij} \sum_{r,s=1}^N p_{n,i}^{(r)} Z_{nm}^{rs} p_{m,j}^{(s)} \right] |V_N\rangle_{1\dots N}^X, \quad (6.253)$$

where $|\hat{V}_N\rangle_{1\dots N}^X$ and $|V_N\rangle_{1\dots N}^X$ are the overlaps in the background corresponding to the world sheet actions (6.242) and (6.238) respectively, the explicit form of the latter is given in appendix A. Clearly the expression (6.253) satisfies the continuity conditions for the modes of the momenta, and the coefficients Z_{nm}^{rs} are determined so that the continuity conditions for the coordinates are satisfied [Sugino (2000)].

$$\bullet |\hat{I}\rangle^X \equiv |\hat{V}_1\rangle^X$$

For the $N = 1$ case, we consider the identity overlap $|\hat{I}\rangle^X \equiv |\hat{V}_1\rangle^X$. The continuity conditions for the momenta require that

$$P_i(\sigma) + P_i(\pi - \sigma) = \frac{2}{\pi} p_i + \frac{2}{\pi \sqrt{\alpha'}} \sum_{n=2,4,6,\dots} \cos(n\sigma) p_{n,i}$$

should vanish for $0 \leq \sigma \leq \frac{\pi}{2}$, namely,

$$p_i = 0, \quad p_{n,i} = 0 \quad (n = 2, 4, 6, \dots), \quad (6.254)$$

which is satisfied by the overlap in the Neumann case $|I\rangle$. In addition, the conditions for the coordinates are that

$$X^j(\sigma) - X^j(\pi - \sigma) = \frac{2}{\pi} \theta^{jk} \left(\sigma - \frac{\pi}{2}\right) p_k + \quad (6.255)$$

$$4\sqrt{\alpha'} \sum_{n=1,3,5,\dots} G^{jk} \cos(n\sigma) x_{n,k} + 4\sqrt{\alpha'} \sum_{n=2,4,6,\dots} \frac{1}{2\pi\alpha'} \theta^{jk} \sin(n\sigma) \frac{1}{n} p_{n,k},$$

should vanish for $0 \leq \sigma \leq \frac{\pi}{2}$. The first and third lines in the r. h. s. can be put to zero by using (6.254). So what we have to consider is the remaining condition $x_{n,k} = 0$ for $n = 1, 3, 5, \dots$, which however is nothing but the continuity condition for the coordinates in the Neumann case. It can be understood from the point that the second line in (6.255) does not depend on θ^{ij} . Thus it turns out that the continuity conditions in the case of the B -field turned on are satisfied by the identity overlap made in the Neumann case. The solution is [Sugino (2000)]

$$|\hat{I}\rangle^X = |I\rangle^X = \exp \left[-\frac{1}{2} G^{ij} \sum_{n=0}^{\infty} (-1)^n a_{n,i}^\dagger a_{n,j}^\dagger \right] |0\rangle, \quad (6.256)$$

where also the zero modes y_i and p_i are written by using the creation and annihilation operators $a_{0,i}^\dagger$ and $a_{0,i}$ as

$$y_i = \frac{i}{2} \sqrt{2\alpha'} (a_{0,i} - a_{0,i}^\dagger), \quad p_i = \frac{1}{\sqrt{2\alpha'}} (a_{0,i} + a_{0,i}^\dagger).$$

$$\bullet |\hat{V}_2\rangle_{12}^X$$

For the $N = 2$ case, we are to do the same argument as in the $N = 1$ case. The continuity conditions mean that

$$\begin{aligned} P_i^{(1)}(\sigma) + P_i^{(2)}(\pi - \sigma) &= \frac{1}{\pi}(p_i^{(1)} + p_i^{(2)}) + \frac{1}{\pi\sqrt{\alpha'}} \sum_{n=1}^{\infty} \cos(n\sigma)(p_{n,i}^{(1)} + (-1)^n p_{n,i}^{(2)}), \\ X^{(1)j}(\sigma) - X^{(2)j}(\pi - \sigma) &= G^{jk}(y_k^{(1)} - y_k^{(2)}) + \frac{1}{\pi}\theta^{jk}\left(\sigma - \frac{\pi}{2}\right)(p_k^{(1)} + p_k^{(2)}) \\ &\quad + 2\sqrt{\alpha'} \sum_{n=1}^{\infty} \left[G^{jk} \cos(n\sigma)(x_{n,k}^{(1)} - (-1)^n x_{n,k}^{(2)}) \right. \\ &\quad \left. + \frac{1}{2\pi\alpha'} \theta^{jk} \sin(n\sigma) \frac{1}{n}(p_{n,k}^{(1)} + (-1)^n p_{n,k}^{(2)}) \right] \end{aligned}$$

should be zero for $0 \leq \sigma \leq \pi$. It turns out again that the conditions for the modes are identical with those in the Neumann case:

$$\begin{aligned} p_i^{(1)} + p_i^{(2)} &= 0, & p_{n,i}^{(1)} + (-1)^n p_{n,i}^{(2)} &= 0, \\ y_i^{(1)} - y_i^{(2)} &= 0, & x_{n,i}^{(1)} - (-1)^n x_{n,i}^{(2)} &= 0, \end{aligned}$$

for $n \geq 1$. Thus we have the solution [Sugino (2000)]

$$|\hat{V}_2\rangle_{12}^X = |V_2\rangle_{12}^X = \exp \left[-G^{ij} \sum_{n=0}^{\infty} (-1)^n a_{n,i}^{(1)\dagger} a_{n,j}^{(2)\dagger} \right] |0\rangle_{12}. \quad (6.257)$$

• $|\hat{V}_4\rangle_{1234}^X$

We find a solution of the continuity conditions (6.241) in the $N = 4$ case assuming the form

$$|\hat{V}_4\rangle_{1234}^X = \exp \left[\frac{i}{4\pi\alpha} \theta^{ij} \sum_{r,s=1}^4 p_{n,i}^{(r)} Z_{nm}^{rs} p_{m,j}^{(s)} \right] |V_4\rangle_{1234}^X. \quad (6.258)$$

When considering the continuity conditions, it is convenient to employ the Z_4 -Fourier transformed variables:

$$\begin{aligned} Q_1^j(\sigma) &= \frac{1}{2} [iX^{(1)j}(\sigma) - X^{(2)j}(\sigma) - iX^{(3)j}(\sigma) + X^{(4)j}(\sigma)] \equiv Q^j(\sigma), \\ Q_2^j(\sigma) &= \frac{1}{2} [-X^{(1)j}(\sigma) + X^{(2)j}(\sigma) - X^{(3)j}(\sigma) + X^{(4)j}(\sigma)], \\ Q_3^j(\sigma) &= \frac{1}{2} [-iX^{(1)j}(\sigma) - X^{(2)j}(\sigma) + iX^{(3)j}(\sigma) + X^{(4)j}(\sigma)] \equiv \bar{Q}^j(\sigma), \\ Q_4^j(\sigma) &= \frac{1}{2} [X^{(1)j}(\sigma) + X^{(2)j}(\sigma) + X^{(3)j}(\sigma) + X^{(4)j}(\sigma)]. \end{aligned}$$

For the momentum variables we also define the Z_4 -Fourier transformed variables $P_{1,i}(\sigma) (\equiv P_i(\sigma))$, $P_{2,i}(\sigma)$, $P_{3,i}(\sigma) (\equiv \bar{P}_i(\sigma))$ and $P_{4,i}(\sigma)$ by the same combinations of $P_i^{(r)}(\sigma)$'s as the above. These variables have the following mode expansions

$$\begin{aligned} P_{t,i}(\sigma) &= \frac{1}{\pi\sqrt{2\alpha'}} P_{t,0,i} + \frac{1}{\pi\sqrt{\alpha'}} \sum_{n=1}^{\infty} \cos(n\sigma) P_{t,n,i}, \\ Q_t^j(\sigma) &= G^{jk} \sqrt{2\alpha'} Q_{t,0,k} + \frac{1}{\pi} \theta^{jk} \left(\sigma - \frac{\pi}{2} \right) \frac{1}{\sqrt{2\alpha'}} P_{t,0,k} \\ &\quad + \sqrt{2\alpha'} \sum_{n=1}^{\infty} \left[G^{jk} \cos(n\sigma) Q_{t,n,k} + \frac{1}{2\pi\alpha'} \theta^{jk} \sin(n\sigma) \frac{1}{n} P_{t,n,k} \right], \end{aligned} \quad (6.259)$$

where $t = 1, 2, 3, 4$. From now on, we frequently omit the subscript t for the $t = 1$ case, and at the same time we employ the expression with a bar instead of putting the subscript t for the $t = 3$ case.

Using those variables, the continuity conditions are written as

$$\begin{aligned} Q_4^j(\sigma) - Q_4^j(\pi - \sigma) &= 0, \quad P_{4,i}(\sigma) + P_{4,i}(\pi - \sigma) = 0, \\ Q_2^j(\sigma) + Q_2^j(\pi - \sigma) &= 0, \quad P_{2,i}(\sigma) - P_{2,i}(\pi - \sigma) = 0, \\ Q^j(\sigma) - iQ^j(\pi - \sigma) &= 0, \quad P_i(\sigma) + i\bar{P}_i(\pi - \sigma) = 0, \\ \bar{Q}^j(\sigma) + i\bar{Q}^j(\pi - \sigma) &= 0, \quad \bar{P}_i(\sigma) - iP_i(\pi - \sigma) = 0 \end{aligned} \quad (6.260)$$

for $0 \leq \sigma \leq \frac{\pi}{2}$. In terms of the modes, the conditions for the sectors of $t = 2$ and 4 are identical with the Neumann case

$$\begin{aligned}(1 - C)|Q_{4,k})|\hat{V}_4\rangle^X &= (1 + C)|P_{4,k})|\hat{V}_4\rangle^X = 0, \\ (1 + C)|Q_{2,k})|\hat{V}_4\rangle^X &= (1 - C)|P_{2,k})|\hat{V}_4\rangle^X = 0,\end{aligned}$$

which can be seen from the point that the conditions (6.260) for the sectors of $t = 2$ and 4 lead the same relations between the modes as those without the terms containing θ^{jk} . Here we adopted the vector notation for the modes

$$|Q_{t,k}) = \begin{bmatrix} Q_{t,0,k} \\ Q_{t,1,k} \\ \vdots \end{bmatrix}, \quad |P_{t,k}) = \begin{bmatrix} P_{t,0,k} \\ P_{t,1,k} \\ \vdots \end{bmatrix},$$

and C is a matrix such that $(C)_{nm} = (-1)^n \delta_{nm}$ ($n, m \geq 0$). Thus there is needed no correction containing θ^{ij} for the sectors of $t = 2$ and 4, so it is natural to assume the form of the phase factor in (6.258) as

$$\frac{1}{2}\theta^{ij} \sum_{r,s=1}^4 (p_i^{(r)}|Z^{rs}|p_j^{(s)}) = \theta^{ij} (P_i|Z|\bar{P}_j) \quad (6.261)$$

with Z being anti-Hermitian.

Next let us consider the conditions for the sectors of $t = 1$ and 3. We rewrite the mode expansions of $Q^j(\sigma)$ and $\bar{Q}^j(\sigma)$ as [Sugino (2000)]

$$\begin{aligned}Q^j(\sigma) &= G^{jk}(\sqrt{2\alpha'}Q_{0,k} + 2\sqrt{\alpha'} \sum_{n=1}^{\infty} \cos(n\sigma)Q_{n,k}) \\ &+ \theta^{jk} \left[\int_{\pi/2}^{\sigma} d\sigma' P_i(\sigma') + \frac{1}{\pi\sqrt{\alpha'}} \sum_{n=1,3,5,\dots} \frac{1}{n} (-1)^{(n-1)/2} P_{n,k} \right] \\ &\equiv \theta^{jk} \int_{\pi/2}^{\sigma} d\sigma' P_i(\sigma') + \Delta Q^j(\sigma),\end{aligned} \quad (6.262)$$

$$\begin{aligned}\bar{Q}^j(\sigma) &= G^{jk}(\sqrt{2\alpha'}\bar{Q}_{0,k} + 2\sqrt{\alpha'} \sum_{n=1}^{\infty} \cos(n\sigma)\bar{Q}_{n,k}) \\ &+ \theta^{jk} \left[\int_{\pi/2}^{\sigma} d\sigma' \bar{P}_i(\sigma') + \frac{1}{\pi\sqrt{\alpha'}} \sum_{n=1,3,5,\dots} \frac{1}{n} (-1)^{(n-1)/2} \bar{P}_{n,k} \right] \\ &\equiv \theta^{jk} \int_{\pi/2}^{\sigma} d\sigma' \bar{P}_i(\sigma') + \Delta \bar{Q}^j(\sigma).\end{aligned} \quad (6.263)$$

Using the conditions for $P_i(\sigma)$ and $\bar{P}_i(\sigma)$ in (6.260), we can reduce the conditions for $Q^j(\sigma)$ and $\bar{Q}^j(\sigma)$ to those for $\Delta Q^j(\sigma)$ and $\Delta \bar{Q}^j(\sigma)$:

$$\begin{aligned}\Delta Q^j(\sigma) &= \begin{cases} i\Delta Q^j(\pi - \sigma) & (0 \leq \sigma \leq \frac{\pi}{2}) \\ -i\Delta Q^j(\pi - \sigma) & (\frac{\pi}{2} \leq \sigma \leq \pi), \end{cases} \\ \Delta \bar{Q}^j(\sigma) &= \begin{cases} -i\Delta \bar{Q}^j(\pi - \sigma) & (0 \leq \sigma \leq \frac{\pi}{2}) \\ i\Delta \bar{Q}^j(\pi - \sigma) & (\frac{\pi}{2} \leq \sigma \leq \pi). \end{cases}\end{aligned}$$

These formulas are translated to the relations between the modes via the Fourier transformation. The result is expressed in the vector notation as

$$(1 - X)|Q_i\rangle|\hat{V}_4\rangle^X = (1 + X)|\bar{Q}_i\rangle|\hat{V}_4\rangle^X = 0, \quad (6.264)$$

where the vectors $|Q_i\rangle$ and $|\bar{Q}_i\rangle$ stand for

$$\begin{aligned}|Q_i\rangle &= \begin{bmatrix} Q_{0,i} + \frac{i}{4\alpha'} G_{ik} \theta^{kj} \sum_{n=0}^{\infty} X_{0n} P_{n,j} \\ Q_{1,i} \\ Q_{2,i} \\ \vdots \end{bmatrix}, \\ |\bar{Q}_i\rangle &= \begin{bmatrix} \bar{Q}_{0,i} + \frac{i}{4\alpha'} G_{ik} \theta^{kj} \sum_{n=0}^{\infty} X_{0n} \bar{P}_{n,j} \\ \bar{Q}_{1,i} \\ \bar{Q}_{2,i} \\ \vdots \end{bmatrix}.\end{aligned}$$

In (6.264), passing the vectors through the phase factor of the $|\hat{V}_4\rangle$ and using the continuity conditions in the Neumann case

$$\begin{aligned}(1 + X)|P_i\rangle|V_4\rangle^X &= (1 - X)|\bar{P}_i\rangle|V_4\rangle^X = 0, \\ (1 - X)|Q_i\rangle|V_4\rangle^X &= (1 + X)|\bar{Q}_i\rangle|V_4\rangle^X = 0,\end{aligned} \quad (6.265)$$

we get the equations, which the coefficients Z_{nm} 's should satisfy,

$$\begin{aligned}[(1 - X)_{m0} \sum_{n=0}^{\infty} (\bar{Z}_{0n} + i\frac{\pi}{2} \bar{X}_{0n}) P_{n,j} + \sum_{n=1}^{\infty} (1 - X)_{mn} \sum_{n'=0}^{\infty} \bar{Z}_{nn'} P_{n',j}] |V_4\rangle^X &= 0 \\ [(1 + X)_{m0} \sum_{n=0}^{\infty} (Z_{0n} - i\frac{\pi}{2} X_{0n}) \bar{P}_{n,j} + \sum_{n=1}^{\infty} (1 + X)_{mn} \sum_{n'=0}^{\infty} Z_{nn'} \bar{P}_{n',j}] |V_4\rangle^X &= 0\end{aligned}$$

for $m \geq 0$. Now all our remaining task is to solve these equations. It is

easy to see that a solution of them is given by [Sugino (2000)]

$$\begin{aligned} Z_{mn} &= -i\frac{\pi}{2}(1-X)_{mn} + i\beta\frac{\pi}{2}C_{mn}, \quad (m, n \geq 0, \text{ except for } m = n = 0), \\ Z_{00} &= i\beta\frac{\pi}{2}, \end{aligned}$$

if we pay attention to (6.265). Here β is an unknown real constant, which is not fixed by the continuity conditions alone. This ambiguity of the solution comes from the property of the matrix X : $XC = -CX$. However it will become clear that the term containing the constant β does not contribute to the vertex $|\hat{V}_4\rangle^X$.

Therefore, we have the expression of the phase (6.261)

$$\begin{aligned} \theta^{ij}(P_i|Z|\bar{P}_j) &= \theta^{ij}\left[i\frac{\pi}{2}P_{0,i}\bar{P}_{0,j} + i\beta\frac{\pi}{2}\sum_{n=0}^{\infty}(-1)^nP_{n,i}\bar{P}_{n,j}\right. \\ &\quad \left.- i\frac{\pi}{2}\sum_{m,n=0}^{\infty}P_{m,i}(1-X)_{mn}\bar{P}_{n,j}\right]. \end{aligned}$$

Then recalling (6.265) again, the last term in the r. h. s. can be discarded. Also we can rewrite the term containing β

$$-\frac{\theta^{ij}}{4\alpha'}\beta(P_i|C|\bar{P}_j) = +\frac{\theta^{ij}}{4\alpha'}\beta(P_i|X^T CX|\bar{P}_j) = +\frac{\theta^{ij}}{4\alpha'}\beta(P_i|C|\bar{P}_j),$$

on $|V_4\rangle^X$. The above formula means that the term containing β can be set to zero on $|V_4\rangle^X$. After all, the form of the 4-string vertex becomes

$$|\hat{V}_4\rangle_{1234}^X = \exp\left[-\frac{\theta^{ij}}{4\alpha'}P_{0,i}\bar{P}_{0,j}\right]|V_4\rangle_{1234}^X.$$

Note that the phase factor has the cyclic symmetric form

$$-\frac{\theta^{ij}}{4\alpha'}P_{0,i}\bar{P}_{0,j} = i\frac{\theta^{ij}}{8\alpha'}(p_{0,i}^{(1)}p_{0,j}^{(2)} + p_{0,i}^{(2)}p_{0,j}^{(3)} + p_{0,i}^{(3)}p_{0,j}^{(4)} + p_{0,i}^{(4)}p_{0,j}^{(1)}),$$

which is a property the vertices should have²².

²²Here the momentum $p_{0,i}^{(r)}$ is given by $p_{0,i}^{(r)} = \sqrt{2\alpha'}p_i^{(r)}$.

$$\bullet |\hat{V}_3\rangle_{123}^X$$

We can get the 3-string overlap in the similar manner as in the 4-string case. First, we introduce the Z_3 -Fourier transformed variables

$$\begin{aligned} Q_1^j(\sigma) &= \frac{1}{\sqrt{3}}[eX^{(1)j}(\sigma) + \bar{e}X^{(2)j}(\sigma) + X^{(3)j}(\sigma)] \equiv Q^j(\sigma), \\ Q_2^j(\sigma) &= \frac{1}{\sqrt{3}}[\bar{e}X^{(1)j}(\sigma) + eX^{(2)j}(\sigma) + X^{(3)j}(\sigma)] \equiv \bar{Q}^j(\sigma), \\ Q_3^j(\sigma) &= \frac{1}{\sqrt{3}}[X^{(1)j}(\sigma) + X^{(2)j}(\sigma) + X^{(3)j}(\sigma)], \end{aligned}$$

where $e \equiv e^{i2\pi/3}$, $\bar{e} \equiv e^{-i2\pi/3}$. The momenta $P_{1,i}(\sigma)(\equiv P_i(\sigma))$, $P_{2,i}(\sigma)(\equiv \bar{P}_i(\sigma))$ and $P_{3,i}(\sigma)$ are defined in the same way. The mode expansions take the same form as those in (6.259). In these variables, the continuity conditions require

$$\begin{aligned} Q^j(\sigma) - eQ^j(\pi - \sigma) &= 0, & P_i(\sigma) + eP_i(\pi - \sigma) &= 0, \\ \bar{Q}^j(\sigma) - \bar{e}\bar{Q}^j(\pi - \sigma) &= 0, & \bar{P}_i(\sigma) + \bar{e}\bar{P}_i(\pi - \sigma) &= 0, \\ Q_3^j(\sigma) - Q_3^j(\pi - \sigma) &= 0, & P_{3,i}(\sigma) + P_{3,i}(\pi - \sigma) &= 0 \end{aligned}$$

for $0 \leq \sigma \leq \frac{\pi}{2}$. The conditions imposed to the modes with respect to the $t = 3$ component are identical with those in the Neumann case

$$(1 + C)|P_{3,i}|\hat{V}_3\rangle^X = (1 - C)|Q_{3,i}|\hat{V}_3\rangle^X = 0.$$

Thus the $t = 3$ component does not couple with θ^{ij} , so we can find the solution by determining the single anti-Hermitian matrix Z in the phase factor whose form is assumed as [Sugino (2000)]

$$\frac{1}{2}\theta^{ij} \sum_{r,s=1}^3 (p_i^{(r)}|Z^{rs}|p_j^{(s)}) = \theta^{ij}(P_i|Z|\bar{P}_j). \quad (6.266)$$

For the sectors of $t = 1$ and 2, the same argument goes on as in the 4-string case. $Q^j(\sigma)$ and $\bar{Q}^j(\sigma)$ have the mode expansions same as in eqs. (6.262) and (6.263). The conditions we have to consider are

$$\begin{aligned} \Delta Q^j(\sigma) &= \begin{cases} e\Delta Q^j(\pi - \sigma), & (0 \leq \sigma \leq \frac{\pi}{2}) \\ \bar{e}\Delta Q^j(\pi - \sigma), & (\frac{\pi}{2} \leq \sigma \leq \pi), \end{cases} \\ \Delta \bar{Q}^j(\sigma) &= \begin{cases} \bar{e}\Delta \bar{Q}^j(\pi - \sigma), & (0 \leq \sigma \leq \frac{\pi}{2}) \\ e\Delta \bar{Q}^j(\pi - \sigma), & (\frac{\pi}{2} \leq \sigma \leq \pi), \end{cases} \end{aligned}$$

which are rewritten as the relations between the modes

$$(1 - Y)|\mathcal{Q}_i\rangle|\hat{V}_3\rangle^X = (1 - Y^T)|\bar{\mathcal{Q}}_i\rangle|\hat{V}_3\rangle^X = 0. \quad (6.267)$$

Recalling the conditions in the Neumann case

$$\begin{aligned} (1 + Y)|P_i\rangle|V_3\rangle^X &= (1 + Y^T)|\bar{P}_i\rangle|V_3\rangle^X = 0, \\ (1 - Y)|Q_i\rangle|V_3\rangle^X &= (1 - Y^T)|\bar{Q}_i\rangle|V_3\rangle^X = 0, \end{aligned} \quad (6.268)$$

we end up with the following equations

$$\begin{aligned} [(1 - Y)_{m0} \sum_{n=0}^{\infty} (\bar{Z}_{0n} + \frac{\pi}{2} \bar{X}_{0n}) P_{n,j} + \sum_{n=1}^{\infty} (1 - Y)_{mn} \sum_{n'=0}^{\infty} \bar{Z}_{nn'} P_{n',j}] |V_3\rangle^X &= 0, \\ [(1 - Y^T)_{m0} \sum_{n=0}^{\infty} (Z_{0n} - i \frac{\pi}{2} X_{0n}) \bar{P}_{n,j} + \sum_{n=1}^{\infty} (1 - Y^T)_{mn} \sum_{n'=0}^{\infty} Z_{nn'} \bar{P}_{n',j}] |V_3\rangle^X &= 0 \end{aligned}$$

for $m \geq 0$. It can be easily found out that the expression

$$\begin{aligned} Z_{mn} &= -i \frac{\pi}{\sqrt{3}} (1 + Y^T)_{mn} \quad (m, n \geq 0, \text{ except for } m = n = 0), \\ Z_{00} &= 0, \end{aligned}$$

satisfies the above equations. It should be noted that in this case, because of $CYC = \bar{Y} \neq -Y$, it does not contain any unknown constant differently from the 4-string case.

Owing to the condition (6.268) we can write the phase factor only in terms of the zero-modes. Finally we have [Sugino (2000)]

$$\begin{aligned} |\hat{V}_3\rangle_{123}^X &= \exp \left[-\frac{\theta^{ij}}{4\sqrt{3}\alpha'} P_{0,i} \bar{P}_{0,j} \right] |V_3\rangle_{123}^X \\ &= \exp \left[i \frac{\theta^{ij}}{12\alpha'} (p_{0,i}^{(1)} p_{0,j}^{(2)} + p_{0,i}^{(2)} p_{0,j}^{(3)} + p_{0,i}^{(3)} p_{0,j}^{(1)}) \right] |V_3\rangle_{123}^X. \end{aligned} \quad (6.269)$$

It is not clear whether the solutions we have obtained here are unique or not. However we can show that the phase factors are consistent with the relations between the overlaps which they should satisfy,

$${}_3\langle \hat{I} | \hat{V}_3 \rangle_{123} = |\hat{V}_2\rangle_{12}, \quad {}_4\langle \hat{I} | \hat{V}_4 \rangle_{1234} = |\hat{V}_3\rangle_{123}, \quad {}_{34}\langle \hat{V}_2 | |\hat{V}_3\rangle_{123} |\hat{V}_3\rangle_{456} = |\hat{V}_4\rangle_{1256},$$

by using the momentum conservation on the vertices $(p_i^{(1)} + \cdots + p_i^{(N)}) |\hat{V}_N\rangle_{1\dots N}^X = 0$. Furthermore we can see that the phase factors successfully reproduce the Moyal product structures of the correlators among

vertex operators obtained in the perturbative approach to open string theory in the constant B -field background [Seiberg and Witten (1999)]. These facts convince us that the solutions obtained here are physically meaningful.

6.5.9.4 Transformation of String Fields

In the previous section, we have explicitly constructed the overlap vertices in the operator formulation under the constant B -field background. Then we have obtained the vertices with a new noncommutative structure of the Moyal type originating from the constant B -field, in addition to the ordinary product $*$ of string fields. Denoting the product with the new structure by \star , the action of the string field theory is written as

$$\begin{aligned} S_B &= \frac{1}{G_s} \int \left(\frac{1}{2} \psi \star Q\psi + \frac{1}{3} \psi \star \psi \star \psi \right) \\ &= \frac{1}{G_s} \left(\frac{1}{2_{12}} \langle \hat{V}_2 | \psi \rangle_1 Q | \psi \rangle_2 + \frac{1}{3_{123}} \langle \hat{V}_3 | \psi \rangle_1 | \psi \rangle_2 | \psi \rangle_3 \right), \quad (6.270) \end{aligned}$$

where the BRST charge Q is constructed from the world sheet action (6.242). The theory (6.270) gives the noncommutative $U(1)$ Yang–Mills theory in the low energy region in the same sense as Witten’s open string field theory in the case of the Neumann boundary condition leads to the ordinary $U(1)$ Yang–Mills theory in the low energy limit.²³

In [Seiberg and Witten (1999)] the authors argued that open string theory in the constant B -field background leads to either commutative or noncommutative Yang–Mills theories, corresponding to the different regularization scheme (the so-called Pauli–Villars regularization or the point-splitting regularization) in the world sheet formulation. They discussed a map between the gauge fields in the commutative and noncommutative Yang–Mills theories. In string field theory perspective, there also should be a certain transformation (hopefully simpler than the Yang–Mills case) from the string field ψ in (6.270) to a string field in a new string field theory which leads to the commutative Yang–Mills theory in the low energy limit.

Here we get the new string field theory by finding a unitary transformation which absorbs the noncommutative structure of the Moyal type in the product \star into a redefinition of the string fields. There are used the two vertices $|\hat{V}_2\rangle$ and $|\hat{V}_3\rangle$ in the action (6.270). Recall that the 2-string vertex is in the same form as in the Neumann case and has no Moyal type

²³It can be explicitly seen by repeating a similar calculation as that carried out in [Dearnaley (1990)].

noncommutative structure. First, we consider the phase factor of the 3-string vertex which multiplies in front of $|V_3\rangle$ (see (6.269)). Making use of the continuity conditions

$$P_{0,i} = -2 \sum_{n=1}^{\infty} Y_{0n} P_{n,i}, \quad \bar{P}_{0,i} = -2 \sum_{n=1}^{\infty} \bar{Y}_{0n} \bar{P}_{n,i}, \quad (6.271)$$

it can be rewritten as [Sugino (2000)]

$$\begin{aligned} -\frac{\theta^{ij}}{4\sqrt{3}\alpha'} P_{0,i} \bar{P}_{0,j} &= \frac{\theta^{ij}}{4\sqrt{3}\alpha'} \sum_{n=1}^{\infty} (P_{0,i} \bar{Y}_{0n} \bar{P}_{n,j} + P_{n,i} Y_{0n} \bar{P}_{0,j}) \\ &= -\frac{\theta^{ij}}{24\alpha'} \sum_{n=1}^{\infty} X_{0n} [(-p_{0,i}^{(2)} - p_{0,i}^{(3)} + 2p_{0,i}^{(1)}) p_{n,j}^{(1)} \\ &\quad + (-p_{0,i}^{(3)} - p_{0,i}^{(1)} + 2p_{0,i}^{(2)}) p_{n,j}^{(2)} + (-p_{0,i}^{(1)} - p_{0,i}^{(2)} + 2p_{0,i}^{(3)}) p_{n,j}^{(3)}] \\ &= -\frac{\theta^{ij}}{8\alpha'} \sum_{r=1}^3 \sum_{n=1}^{\infty} X_{0n} p_{0,i}^{(r)} p_{n,j}^{(r)}, \end{aligned}$$

where we used the property of the matrix Y : $Y_{0n} = -\bar{Y}_{0n} = \frac{\sqrt{3}}{2} X_{0n}$ for $n \geq 1$ and the momentum conservation on $|V_3\rangle$: $p_{0,i}^{(1)} + p_{0,i}^{(2)} + p_{0,i}^{(3)} = 0$. We manage to represent the phase factor of the Moyal type as a form factorized into the product of the unitary operators

$$\mathcal{U}_r = \exp \left(\frac{\theta^{ij}}{8\alpha'} \sum_{n=1,3,5,\dots} X_{0n} p_{0,i}^{(r)} p_{n,j}^{(r)} \right). \quad (6.272)$$

Note that the unitary operator acts to a single string field. So the Moyal type noncommutativity can be absorbed by the unitary rotation of the string field

$${}_{123} \langle \hat{V}_3 | |\psi\rangle_1 |\psi\rangle_2 |\psi\rangle_3 = {}_{123} \langle V_3 | \mathcal{U}_1 \mathcal{U}_2 \mathcal{U}_3 | \psi\rangle_1 |\psi\rangle_2 |\psi\rangle_3 = {}_{123} \langle V_3 | |\tilde{\psi}\rangle_1 |\tilde{\psi}\rangle_2 |\tilde{\psi}\rangle_3, \quad (6.273)$$

with $|\tilde{\psi}\rangle_r = \mathcal{U}_r |\psi\rangle_r$. It should be remarked that this manipulation has been succeeded owing to the factorized expression of the phase factor, which originates from the continuity conditions relating the zero-modes to the nonzero-modes (6.271). It is a characteristic feature of string field theory that can not be found in any local field theories.

Next let us see the kinetic term. In doing so, it is judicious to write the kinetic term as follows:

$${}_{12} \langle \hat{V}_2 | |\psi\rangle_1 (Q|\psi\rangle_2) = {}_{123} \langle \hat{V}_3 | |\psi\rangle_1 (Q_L |I\rangle_2 |\psi\rangle_3 + |\psi\rangle_2 Q_L |I\rangle_3), \quad (6.274)$$

where Q_L is defined by integrating the BRST current $j_{BRST}(\sigma)$ with respect to σ over the left half region

$$Q_L = \int_0^{\pi/2} d\sigma j_{BRST}(\sigma).$$

Equation (6.274) is also represented by the product \star as

$$\psi \star (Q\psi) = \psi \star [(Q_L I) \star \psi + \psi \star (Q_L I)]. \quad (6.275)$$

Here, I stands for the identity element with respect to the \star -product, carrying the ghost number $-\frac{3}{2}$, which corresponds to $|I\rangle$ in the operator formulation. As is discussed by [Horowitz *et. al.* (1986)], in order to show the relation (6.275) we need the formulas

$$Q_R I = -Q_L I, \quad (Q_R \psi) \star \xi = -(-1)^{n_\psi} \psi \star (Q_L \xi) \quad (6.276)$$

for arbitrary string fields ψ and ξ , where Q_R is the integrated BRST current over the right half region of σ . n_ψ stands for the ghost number of the string field ψ minus $\frac{1}{2}$, and takes an integer value. The first formula means that the identity element is a physical quantity, also the second does the conservation of the BRST charge. By using these formulas, the first term in the bracket in r. h. s. of (6.275) becomes

$$(Q_L I) \star \psi = -(Q_R I) \star \psi = I \star (Q_L \psi) = Q_L \psi.$$

Also, it turns out that the second term is equal to $Q_R \psi$. Combining these, we can see that (6.275) holds.

Further, we should remark that because the BRST current does not contain the center of mass coordinate x^j , it commute with the momentum p_i . From the continuity condition $p_i |I\rangle = 0$, it can be seen that $p_i Q_L |I\rangle = 0$. Expanding the exponential in the expression of the unitary operator (6.272) and passing the momentum $p_{0,i}$ to the right, we get

$$\mathcal{U} Q_L |I\rangle = Q_L |I\rangle. \quad (6.277)$$

Now we can write down the result of the kinetic term. As a result of the

same manipulation as in eq. (6.273) and the use of eq. (6.277), we have²⁴

$$\begin{aligned} {}_{12}\langle \hat{V}_2 || \psi \rangle_1 (Q | \psi \rangle_2) &= {}_{123} \langle \hat{V}_3 || \psi \rangle_1 (Q_L | I \rangle_2 | \psi \rangle_3 + | \psi \rangle_2 Q_L | I \rangle_3) \\ &= {}_{123} \langle V_3 || \tilde{\psi} \rangle_1 (Q_L | I \rangle_2 | \tilde{\psi} \rangle_3 + | \tilde{\psi} \rangle_2 Q_L | I \rangle_3) = {}_{12} \langle V_2 || \tilde{\psi} \rangle_1 (Q | \tilde{\psi} \rangle_2). \end{aligned} \quad (6.278)$$

Here we have a comment [Sugino (2000)]. If we considered the kinetic term itself without using (6.274), what would be going on? Let us see this. From the continuity conditions for $|\hat{V}_2\rangle_{12}^X = |V_2\rangle_{12}^X$:

$$p_{0,i}^{(1)} + p_{0,i}^{(2)} = 0, \quad p_{n,i}^{(1)} + (-1)^n p_{n,i}^{(2)} = 0 \quad (n = 1, 2, \dots),$$

it could be shown that the 2-string overlap is invariant under the unitary rotation

$$\mathcal{U}_1 \mathcal{U}_2 |V_2\rangle_{12} = |V_2\rangle_{12}.$$

So we would find the expression for the kinetic term after the rotation

$${}_{12} \langle V_2 || \psi \rangle_1 Q | \psi \rangle_2 = {}_{12} \langle V_2 || \tilde{\psi} \rangle_1 \tilde{Q} | \tilde{\psi} \rangle_2,$$

where \tilde{Q} is the BRST charge similarity transformed by \mathcal{U}

$$\tilde{Q} = \mathcal{U} Q \mathcal{U}^\dagger. \quad (6.279)$$

However, after some computations of the r. h. s. of (6.279), we could see that \tilde{Q} has divergent term proportional to

$$\sum_{n=1,3,5,\dots} 1$$

and thus it is not well-defined. It seems that this procedure is not correct and needs some suitable regularization, which preserves the conformal symmetry²⁵. It is considered that the use of eq. (6.274) gives that kind of

²⁴Strictly speaking, in general this formula holds in the case that both of the string fields $|\psi\rangle$ and $|\tilde{\psi}\rangle$ belong to the Fock space which consists of states excited by *finite* number of creation operators. This point is subtle for giving a proof. However, for the infinitesimal θ case, by keeping arbitrary finite order terms in the expanded form of the exponential of \mathcal{U} , we can make the situation of both $|\psi\rangle$ and $|\tilde{\psi}\rangle$ being inside the Fock space, and thus clearly eq. (6.278) holds. From this fact, it is plausible to expect that eq. (6.278) is correct in the finite θ case.

²⁵That divergence comes from the mid-point singularity of the string coordinates transformed by \mathcal{U} . In fact, after some calculations, we have

$$\mathcal{U} X^j(\sigma) \mathcal{U}^\dagger = X^j(\sigma) - i \frac{\theta^{jk}}{4\sqrt{2}\alpha'} \sum_{n=1,3,5,\dots} X_{n0} p_{n,k} - \frac{\theta^{jk}}{4} p_k \operatorname{sgn} \left(\sigma - \frac{\pi}{2} \right). \quad (6.280)$$

The last term leads to the mid-point singularity in the energy-momentum tensor and

regularization, which will be justified at the end of the next section.

Therefore, the string field theory action (6.270) with the Moyal type noncommutativity added to the ordinary noncommutativity is equivalently rewritten as the one with the ordinary noncommutativity alone [Sugino (2000)]:

$$\begin{aligned} S_B &= \frac{1}{G_s} \int \left(\frac{1}{2} \tilde{\psi} * Q \tilde{\psi} + \frac{1}{3} \tilde{\psi} * \tilde{\psi} * \tilde{\psi} \right) \\ &= \frac{1}{G_s} \left(\frac{1}{2} {}_{12} \langle V_2 | |\tilde{\psi}\rangle_1 Q |\tilde{\psi}\rangle_2 + \frac{1}{3} {}_{123} \langle V_3 | |\tilde{\psi}\rangle_1 |\tilde{\psi}\rangle_2 |\tilde{\psi}\rangle_3 \right). \end{aligned} \quad (6.281)$$

It is noted that the BRST charge Q , which is constructed from the world sheet action (6.242), has the same form as the one obtained from the action (6.238) with the relation (6.247). So all the B -dependence has been stuffed into the string fields except that existing in the metric G_{ij} . Furthermore, recalling that the relation between the metrics G^{ij} and g_{ij} is the same form as the T-duality inversion transformation, which was pointed out at the end of section 3, we can make the metric g_{ij} appear in the overlap vertices, instead of the metric G_{ij} . To do so, we consider the following transformation for the modes:

$$\hat{\alpha}_n^i = (E^{T-1})^{ik} \alpha_{n,k}, \quad \hat{p}^i = (E^{T-1})^{ik} p_k, \quad \hat{x}_i = E_{ik} x^k. \quad (6.282)$$

By this transformation, the commutators become

$$[\hat{\alpha}_n^i, \hat{\alpha}_m^j] = n g^{ij} \delta_{n+m,0}, \quad [\hat{p}^i, \hat{x}_j] = -i \delta_j^i,$$

and the bilinear form of the modes

$$G^{ij} \alpha_{n,i} \alpha_{m,j} = g_{ij} \hat{\alpha}_n^i \hat{\alpha}_m^j, \quad G^{ij} p_i \alpha_{m,j} = g_{ij} \hat{p}^i \hat{\alpha}_m^j, \quad G^{ij} p_i p_j = g_{ij} \hat{p}^i \hat{p}^j. \quad (6.283)$$

6.6 Application: Dynamics of Strings and Branes

Dynamics of Nambu–Goto strings and branes was analyzed recently in [Golovnev and Prokhorov (2005)]. It was shown that they could be con-

the BRST charge Q . It seems that the use of (6.274) corresponds to taking the point splitting regularization with respect to the mid-point. Because of the discontinuity of the last term in (6.280), it is considered that the transformed string coordinates have no longer a good picture as a string. It could be understood from the point that the transformation \mathcal{U} drives states around a perturbative vacuum to those around highly non-perturbative one like coherent states.

sidered as continuous limits of ordered discrete sets of relativistic particles for which the tangential velocities were excluded from the action. The authors have proved that p -branes might be considered as continuous limits of discrete sets of relativistic particles and that the p -brane action

$$S = -\gamma \int d^{p+1}\sigma \sqrt{(-1)^p g}, \quad g = \det \frac{\partial x^\mu}{\partial \sigma_a} \frac{\partial x_\mu}{\partial \sigma_b}, \quad (6.284)$$

is a continuous limit of sum of properly modified relativistic particle actions. Here $a, b = 0, 1, \dots, p$ and $\mu = 0, 1, \dots, n$ where $p + 1$ and $n + 1$ are the brane world-sheet and the bulk space dimensions respectively; g is the induced metric determinant on the world-sheet. Strings correspond to the $p = 1$ case in (6.284). The modification is such that particle motions along the brane hypersurface become un-physical. It yields p constraints; the remaining constraint ($H = 0$) is a consequence of arbitrariness of ‘time’ σ_0 . The authors have also found the linear constraints in un-physical momenta, which allowed them to derive the evolution operators for the objects under consideration from the ‘first principles’.

6.6.1 A Relativistic Particle

Following [Golovnev and Prokhorov (2005)], we start with the simplest case of one relativistic particle which can be formally considered as a 0-brane. Recall that the motion of free relativistic particle is defined by the well-known action

$$S = -m \int \sqrt{1 - \vec{v}^2} dt,$$

where $\vec{v} = \frac{d\vec{x}(t)}{dt}$. The canonical momentum is

$$\vec{p} \equiv \frac{\partial L}{\partial \vec{v}} = \frac{m \vec{v}}{\sqrt{1 - \vec{v}^2}} \equiv E_p \vec{v},$$

with L being the Lagrangian, and the Hamiltonian is

$$H = E_p = \sqrt{m^2 + \vec{p}^2}.$$

We can write down the action in the explicitly relativistic invariant form by parametrization of the world line: $x^\mu = x^\mu(\sigma_0)$ (usually one uses τ instead of σ_0) with

$$x^\mu(\sigma_0) = (t(\sigma_0), \vec{x}(\sigma_0)), \quad (\mu = 0, \dots, n).$$

We denote

$$\dot{x}^\mu \equiv \frac{dx^\mu}{d\sigma_0}, \quad \text{so that} \quad \vec{v} = \frac{\dot{x}}{\dot{x}^0} \frac{d\sigma_0}{dt}, \quad \text{and}$$

$$S = -m \int \sqrt{\dot{x}^\mu \dot{x}_\mu} d\sigma_0.$$

It is the $p = 0$ case of (6.284) in which we put m (particle's mass) instead of γ . The vector of canonical momentum is

$$p_\mu \equiv \frac{\partial L}{\partial \dot{x}^\mu} = -m \frac{\dot{x}_\mu}{\sqrt{\dot{x}^2}}. \quad (6.285)$$

The obtained theory is invariant under reparametrization group $\sigma_0 \rightarrow \tilde{\sigma}_0 = f(\sigma_0)$, hence its Hamiltonian is zero and by squaring the equation (6.285) one gets a constraint:

$$p^2 - m^2 = 0. \quad (6.286)$$

Note that the information about the sign of p_0 is lost. On the other hand one can prove that any constraint has to be linear in un-physical momenta [Golovnev and Prokhorov (2005)]. One can get the solution in the following way.

From the constraint (6.286) one finds

$$p_0 = \pm \sqrt{m^2 + \vec{p}^2},$$

and it is obvious from (6.285) that $\text{sgn} p_0 = -\text{sgn} \dot{x}_0$. Combining these facts we have

$$p_0 + E_p \text{sgn}(\dot{x}^0) = 0, \quad (6.287)$$

$$\text{with} \quad E_p(\vec{p}) = \sqrt{m^2 + \vec{p}^2}.$$

The Hamiltonian is zero and the total *Dirac Hamiltonian* is [Dirac (1982)]

$$H_T = v (p_0 + E_p \text{sgn}(\dot{x}^0)).$$

Here v is the Lagrange multiplier. Strictly speaking (6.287) is not a constraint because it contains velocity (and H_T is not a Hamiltonian due to the same reason). But it depends only upon the sign of \dot{x} , and this fact allows us to formulate the quantum theory.

We fix the σ_0 'time arrow' by condition

$$\frac{\partial x^0}{\partial \sigma_0} > 0, \quad (6.288)$$

which forces us to admit that $v > 0$, as

$$\Delta x^0 = v \Delta \sigma_0.$$

We use the evolution operator

$$U_\omega(x, \tilde{x}) = \langle x | \exp(-i\omega \hat{H}_T) | \tilde{x} \rangle,$$

$$\psi(x, \sigma_0 + \omega) = \int d^4 \tilde{x} U_\omega(x, \tilde{x}) \psi(\tilde{x}, \sigma_0)$$

and the following relation for infinitesimal ω :

$$\langle x | \exp(-i\omega \hat{H}_T) | \tilde{x} \rangle = \int d^4 p \langle x | \exp(-i\omega H_T(x, p)) | p \rangle \langle p | \tilde{x} \rangle.$$

Integrating over p_0 and (with use of δ -function $\delta(x^0 - \tilde{x}^0 - \omega v)$) over \tilde{x}_0 we get for the wave function [Golovnev and Prokhorov (2005)]:

$$\psi(x^0, \vec{x}) = \int \frac{d^3 p d^3 \tilde{x}}{(2\pi)^3} \exp(i[p_i \Delta x^i - \Delta x^0 E_p]) \psi(\tilde{x}^0, \vec{\tilde{x}}), \quad (6.289)$$

where $\Delta x^0 = v\omega$, and we omit the argument σ_0 because all the information on σ_0 is accumulated in x^0 . Using (6.289) one can get the right Feynman propagator for a relativistic particle.

6.6.2 A String

Now we show that any string and brane can be described as a system of particles. More precisely, the action (6.284) may be regarded as a continuous limit of sum of free relativistic particle actions provided that we take \vec{v}_\perp instead of \vec{v} , where \vec{v}_\perp is the part of velocity orthogonal to the constant time hypersurface of the *brane world-sheet*. We deal here with a kind of indirectly introduced particle interaction.

In this section we consider a string (1-brane) and reproduce the proof of our statement by an explicit calculation [Barbashov and Nesterenko (1990)]. We consider $N + 1$ particles with the position vectors $\vec{x}_k(x^0)$, ($k = 0, 1, \dots, N$) and the action

$$S = -m \sum_{k=0}^N \int dx^0 \sqrt{1 - \vec{v}_{k\perp}^2(x^0)}, \quad \text{in which} \quad \vec{v}_k = \frac{d\vec{x}_k}{dx^0}, \quad (6.290)$$

and

$$\vec{v}_{k\perp} = \vec{v}_k - \frac{(\vec{v}_k \cdot \Delta \vec{x}_k)}{(\Delta \vec{x}_k)^2} \Delta \vec{x}_k, \quad \text{with} \quad \Delta \vec{x}_k \equiv \vec{x}_{k+1} - \vec{x}_k.$$

In the continuous limit we define

$$\frac{kA}{N} \rightarrow \sigma_1, \quad \frac{\Delta \vec{x}_k}{A/N} \rightarrow \vec{k}(\sigma_1), \quad \frac{m}{|\Delta \vec{x}_k|} \rightarrow \gamma.$$

Here $\sigma_1 \in [0, A]$; usually one takes $A = \pi$, but for our purposes it may be natural to consider A as the string length and

$$|\Delta \vec{x}_k| = \frac{A}{N}, \quad |\vec{k}| = 1.$$

In any case we have

$$S = -\gamma \sum_{k=0}^N \int dx^0 \Delta l \sqrt{1 - \vec{v}_{k\perp}^2} \rightarrow -\gamma \int dx^0 |\vec{k}| d\sigma_1 \sqrt{1 - \vec{v}_{\perp}^2},$$

with

$$\vec{k} = \frac{\partial \vec{x}(x^0, \sigma_1)}{\partial \sigma_1}, \quad \vec{v} = \frac{\partial \vec{x}(x^0, \sigma_1)}{\partial x^0}, \quad \vec{v}_{\perp} = \vec{v} - \frac{(\vec{v} \cdot \vec{k})}{k^2} \vec{k}.$$

The *string length* is equal to

$$L = \int |\vec{k}| d\sigma_1.$$

After that we parameterize the world-sheet

$$x^0 = x^0(\sigma_0, \sigma_1), \quad \vec{x} = \vec{x}(\sigma_0, \sigma_1)$$

by introducing a new parameter σ_0 (again the standard notations are $x^0 \rightsquigarrow t$ and $\sigma_0 \rightsquigarrow \tau$). We get

$$\dot{x} \equiv \frac{\partial x(\sigma_0, \sigma_1)}{\partial \sigma_0} = (1, \vec{v}) \dot{x}_0, \quad x' \equiv \frac{\partial x(\sigma_0, \sigma_1)}{\partial \sigma_1} = (x_0', \vec{k} + \vec{v} x_0'),$$

and
$$S = -\gamma \int d^2 \sigma \dot{x}^0 |\vec{k}| \sqrt{1 - \vec{v}_{\perp}^2}.$$

The last expression is equal to the *Nambu-Goto action*:

$$S = -\gamma \int d^2 \sigma \sqrt{(\dot{x} x')^2 - \dot{x}^2 x'^2}.$$

Note that one could start with it and get

$$S = -\gamma \int dx^0 dl \sqrt{1 - \vec{v}_{\perp}^2},$$

which is a continuous limit of (6.290).

We can also propose another discrete analogue of the Nambu–Goto string. Let's consider the following action:

$$S = -m \sum_{k=0}^N \int d\sigma_0 \sqrt{\dot{x}_{k\perp}^\mu \dot{x}_{\mu k\perp}},$$

where $\dot{x}_{k\perp}^\mu$ is the part of \dot{x}_k^μ perpendicular to $x_{k+1}^\mu - x_k^\mu$. The continuous limit is

$$N \rightarrow \infty, \quad \frac{kA}{N} \rightarrow \sigma_1, \quad \frac{m}{|\Delta s_k|} \rightarrow \gamma,$$

with the *invariant interval*

$$(\Delta s_k)^2 = (x_{k+1}^\mu - x_k^\mu)(x_{\mu k+1} - x_{\mu k}).$$

We have

$$\begin{aligned} \dot{x}_\perp^\mu &= \dot{x}^\mu - \frac{\dot{x}^\nu x'_\nu}{x'^\rho x'_\rho} x'^\mu, & \left(\frac{ds}{d\sigma_1} \right)^2 &= x'^2, & \text{and} \\ S &= -\gamma \int d\tau |ds| \sqrt{\dot{x}^2 - \frac{(\dot{x}x')^2}{x'^2}} = -\gamma \int d\sigma_0 d\sigma_1 \sqrt{(\dot{x}x')^2 - \dot{x}^2 x'^2}. \end{aligned}$$

In contrast to the previous paragraph the presented discrete theory has the relativistic invariant form from the very beginning but even the sense of \dot{x}_\perp depends upon the parametrization of the world-sheet. In the gauge $\sigma_0 = x^0$ these two approaches coincide.

6.6.3 A Brane

Now we shall prove our statement for $p > 1$. We consider $(N+1)^p$ particles arranged into some p D lattice with the position vectors $\vec{x}_{i_1 i_2 \dots i_p}$ and the action [Golovnev and Prokhorov (2005)]

$$S = -m \int dx^0 \sum_{i_1=0}^N \dots \sum_{i_p=0}^N \sqrt{1 - \vec{v}_{i_1, \dots, i_p \perp}^2},$$

where $\vec{v}_{i_1, \dots, i_p \perp}$ is the component of $\vec{v}_{i_1 \dots i_p}$ perpendicular to $\vec{x}_{i_1 \dots i_k + 1 \dots i_p} - \vec{x}_{i_1 \dots i_k \dots i_p}$ for all k . In continuous limit we demand $\frac{A i_k}{N} \rightarrow \sigma_k$ and $\frac{m}{\Delta V} \rightarrow \gamma$ with ΔV being volume of a cell of the lattice. The action takes the form

$$S = -\gamma \int dx_0 dV \sqrt{1 - \vec{v}_\perp^2},$$

and we need to prove that S is equal to (6.284).

For the sake of simplicity first we consider some special coordinate system on the brane. Let the brane be parametrized by $\sigma_0, \sigma_1, \dots, \sigma_p$, $\sigma_i \in [0, A]$ for $i = 1, \dots, p$. We choose a coordinate system in which $\sigma_0 = x^0$, so that $\frac{\partial x^0}{\partial \sigma_i} = 0$, and $\frac{\partial x^\mu}{\partial \sigma_i} \frac{\partial x_\mu}{\partial \sigma_j} = 0$, $i \neq j$. It is always possible, locally at least. Let's denote

$$\vec{k}_i \equiv \frac{\partial \vec{x}(x^0, \sigma_i)}{\partial \sigma_i} \quad \text{and} \quad \vec{v} \equiv \frac{\partial \vec{x}(x^0, \sigma_i)}{\partial x^0}$$

on the world-sheet. In our coordinate system \vec{k}_i is orthogonal to \vec{k}_j , $i \neq j$ and

$$\vec{v}_\perp = \vec{v} - \sum_{i=1}^p \frac{(\vec{v} \cdot \vec{k}_i)}{k_i^2} \vec{k}_i.$$

Note that one can find \vec{v}_\perp without the orthogonality condition with the use of standard orthogonalization procedure.

We get

$$\frac{\partial x^\mu}{\partial \sigma_0} = (1, \vec{v}) \quad \text{and} \quad \frac{\partial x^\mu}{\partial \sigma_i} = (0, \vec{k}_i),$$

so the determinant in (6.284) is

$$\begin{aligned} \det \frac{\partial x^\mu}{\partial \sigma_a} \frac{\partial x_\mu}{\partial \sigma_b} &= \begin{vmatrix} 1 - \vec{v}^2 & -\vec{v} \cdot \vec{k}_1 & -\vec{v} \cdot \vec{k}_2 & \dots & -\vec{v} \cdot \vec{k}_p \\ -\vec{v} \cdot \vec{k}_1 & -k_1^2 & 0 & \dots & 0 \\ -\vec{v} \cdot \vec{k}_2 & 0 & -k_2^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\vec{v} \cdot \vec{k}_p & 0 & 0 & \dots & -k_p^2 \end{vmatrix} \\ &= \left(\prod_{i=1}^p (-k_i^2) \right) \left(1 - \vec{v}^2 + \sum_{i=1}^p \frac{(\vec{v} \cdot \vec{k}_i)^2}{k_i^2} \right) \end{aligned}$$

where $a, b = 0, 1, \dots, p$. It is not also difficult to see that the *volume element* at the constant time hypersurface on the brane world-sheet is

$$dV = \prod_{i=1}^p |k_i| d\sigma_i.$$

Thus, we can conclude that the action is

$$S = -\gamma \int d\sigma_0 d^n \sigma_i \sqrt{\left| \det \frac{\partial x^\mu}{\partial \sigma_a} \frac{\partial x_\mu}{\partial \sigma_b} \right|} = -\gamma \int dx_0 dV \sqrt{1 - \vec{v}_\perp^2}. \quad (6.291)$$

6.6.4 String Dynamics

As it was mentioned above, the evolution operator U_ω can be constructed for strings and branes in the same way as above [Goto (1971)]. For $p = 1$ action (6.284) is the action of free bosonic string with the *Lagrangian density* (see, e.g., [Kaku (1988)])

$$\mathcal{L} = -\gamma \sqrt{-g} = -\gamma \sqrt{(\dot{x}x')^2 - \dot{x}^2 x'^2}. \quad (6.292)$$

We assume that \dot{x} is time-like and x' is space-like, so that σ_0 can be regarded as a time parameter. In this case the momentum

$$p_\mu \equiv \frac{\partial \mathcal{L}}{\partial \dot{x}^\mu} = -\gamma \frac{(\dot{x}x')x'_\mu - x'^2 \dot{x}_\mu}{\sqrt{(\dot{x}x')^2 - \dot{x}^2 x'^2}} \quad (6.293)$$

will also be time-like. One easily gets two constraints

$$p_\mu x'^\mu = 0, \quad (6.294)$$

$$p^2 + \gamma^2 x'^2 = 0. \quad (6.295)$$

The second one is obtained by squaring the (6.293) and hence some information is lost. As in the case of a pointlike particle (6.295) yields $p_0 = \pm E_p$ with

$$E_p = \sqrt{\vec{p}^2 - \gamma^2 x'^2}.$$

The sign of p_0 follows from the definition of the momentum. We have

$$p_0 + E_p(x, \vec{p}) \operatorname{sgn}(y_0) = 0, \quad \text{where} \\ y_\mu = (\dot{x}x')x'_\mu - x'^2 \dot{x}_\mu.$$

The vector y is obviously time-like (indeed, $y^2 = x'^2 g > 0$ and $y_\mu \dot{x}^\mu = -g > 0$, so that $\operatorname{sgn} \dot{x}^0 = \operatorname{sgn} y^0$). We get the ‘constraint’ analogous to (6.287):

$$p_0 + E_p(x, \vec{p}) \operatorname{sgn}(\dot{x}^0) = 0. \quad (6.296)$$

The ‘Hamiltonian’ of the theory,

$$H = p_0 + E_p \operatorname{sgn} \dot{x}^0,$$

is zero, and demanding again $\operatorname{sgn} \dot{x}^0 > 0$, we get the *total Hamiltonian*

$$H_T = u(p_0 + E_p) + v p_\mu x'^\mu.$$

We have two un-physical momenta and need to exclude them from E_p . Let’s exclude p_0 and p_1 (we denote the remaining components by the lower index ‘ $>$ ’: $p^\mu = (p_0, p_1, p_{>})$). One can find p_1 from (6.294) if $x'_1 \neq 0$. Above we restricted ourselves to the case $x'_0 = 0$ and had

$$E_p(x, p_{>}) = \sqrt{\left(\frac{p_{>} x'_{>}}{x'_1}\right)^2 + p_{>}^2 - \gamma^2 x'^2}.$$

In general case one can substitute p_1 from (6.294) to (6.295) and get a quadratic equation for p_0 :

$$p_0^2 \left(1 - \left(\frac{x'^0}{x'^1}\right)^2\right) + 2p_0 \frac{(p_{>} x'_{>}) x'^0}{(x'^1)^2} - p_{>}^2 \left(1 + \left(\frac{x'_{>}}{x'^1}\right)^2\right) + \gamma^2 x'^2 = 0.$$

If $|x'^1| > |x'^0|$, it has two real roots of opposite signs, and we can choose a proper one following (6.296). Otherwise we have to try to exclude another component of p_μ from E_p . It’s always possible because x' is space-like and $|\vec{x}'| > |x'^0|$. Above we have seen that the un-physical degree of freedom is related to the motion of particles along the string.

Now we write down the evolution equation [Golovnev and Prokhorov (2005)]

$$\begin{aligned} \psi(x) &= \int \mathcal{D}^{n+1} \tilde{x}(\sigma_1) \left\langle x(\sigma_1) \left| \exp(-i\omega \hat{H}_T) \right| \tilde{x}(\sigma_1) \right\rangle \psi(\tilde{x}(\sigma_1)) = \\ &= \int \mathcal{D}^{n+1} p \mathcal{D}^{n+1} \tilde{x} \exp(i[p_\mu \Delta x^\mu - \omega u(p_0 + E_p(\tilde{x}, p_\perp)) - \omega v p_\mu x'^\mu]) \psi(\tilde{x}) = \\ &= \int \mathcal{D}^{n-1} p_{>} \mathcal{D}^{n+1} \tilde{x} \exp(i[-p_{>} \Delta x_{>} - \omega u E_p(\tilde{x}, p_{>}) + \omega v p_{>} x'_{>}]) \times \\ &\quad \times \delta(\Delta x^0 - \omega u - \omega v x'^0) \delta(-\Delta x_1 + \omega v x'_1) \psi(\tilde{x}). \end{aligned}$$

Here $\mathcal{D}\tilde{x}$ and $\mathcal{D}p$ denote differentials in the functional spaces and all the integrals are path integrals.

δ -Functions determine the Lagrange multipliers

$$\omega v = \frac{\Delta x_1}{x'_1} \quad \text{and} \quad \omega u = \Delta x^0 - \frac{x'^0 \Delta x_1}{x'_1},$$

giving the final result

$$\begin{aligned} \psi(x) = \int \mathcal{D}^{n-1} p_{>} \mathcal{D}^{n-1} \tilde{x}_{>} \exp(i[-p_{>} \Delta x_{>} - \\ - \left(\Delta x^0 - \frac{x'^0 \Delta x_1}{x'_1} \right) E_p(\tilde{x}, p_{>}) + \frac{\Delta x_1}{x'_1} p_{>} x'_{>}]) \psi(\tilde{x}). \end{aligned}$$

If $x'^0 = 0$, (6.288) implies $u > 0$.

6.6.5 Brane Dynamics

We turn to the general case of action (6.284). We denote the $\sigma_0, \sigma_1, \dots, \sigma_p$ derivatives of x by $x_{,0}, x_{,1}, \dots, x_{,p}$. Again we assume that the vector $x_{,0}^\mu$ is time-like, and vectors $x_{,i}^\mu$ are space-like (here and hereafter in this chapter $i, k, l = 1, \dots, p$ while $a, b = 0, \dots, p$). Now in the action (6.284)

$$g(\sigma) = \frac{1}{(p+1)!} \epsilon_{a_0 \dots a_p} \epsilon^{b_0 \dots b_p} x_{\alpha_0, b_0} x^{\alpha_0, a_0} \dots x_{\alpha_p, b_p} x^{\alpha_p, a_p},$$

with ϵ being the unit antisymmetric *Levi-Civita symbol*, and the canonical momentum is [Golovnev and Prokhorov (2005)]

$$p_\mu = \frac{-\gamma(-1)^p}{p! \sqrt{(-1)^p g}} \epsilon_{0 a_1 \dots a_p} \epsilon^{b_0 \dots b_p} x_{\mu, b_0} x_{, b_1} \cdot x^{, a_1} \dots x_{, b_p} \cdot x^{, a_p}.$$

Evidently $p_\mu x_{,i}^\mu = 0$ due to antisymmetry of ϵ , and using the equality

$$\begin{aligned} \epsilon^{a_0 \dots a_p} g(\sigma) &= \epsilon^{b_0 \dots b_p} x_{, b_0} \cdot x^{, a_0} \dots x_{, b_p} \cdot x^{, a_p} \quad \text{one gets} \\ p^2 &= (-1)^p \gamma^2 \zeta(x), \quad \text{with} \quad \zeta(x) = \det x_{,i}^\mu x_{\mu, k}. \end{aligned}$$

So, with the loss of information about the sign of p_0 , the constraints are

$$p_\mu x_{,i}^\mu = 0, \quad (i = 1, 2, \dots, p), \quad (6.297)$$

$$p^2 - (-1)^p \gamma^2 \zeta(x) = 0. \quad (6.298)$$

From (6.298) we have $p_0 = \pm E_p$ with

$$E_p = \sqrt{\vec{p}^2 + (-1)^p \gamma^2 \zeta(x)}.$$

Again, the p_0 sign can be easily found: p_μ and \dot{x}_μ are time-like and

$$p_\mu \dot{x}^\mu = -\gamma \sqrt{(-1)^p g} < 0, \quad \text{so that} \quad \text{sgn}(p_0) = -\text{sgn}(\dot{x}_0).$$

The result is similar to (6.287):

$$p_0 + E_p \text{sgn}(\dot{x}^0) = 0,$$

the ‘Hamiltonian’ is equal to zero, and the total Hamiltonian is

$$H_T = u(p_0 + E_p(x, \vec{p})) + v_i p_\mu x_i^\mu.$$

Here, $p+1$ momenta are un-physical ones. We assume that $\det(x_{i,k}) \neq 0$ ($x_{i,k}$ is an $p \times p$ matrix) and exclude momenta p_1, \dots, p_p from E_p . Due to (6.297) we have

$$p_i = ([x_{..}]^{-1})_{il}(p_0 x_{0,l} - p_{>} x_{>,l}).$$

Here we denoted all the components of p^μ with $\mu > p$ by the lower index ‘ $>$ ’, and $([x_{..}]^{-1})_{il} \equiv d_{il}$ is a matrix inverse of $x_{l,i}$. Then (6.298) turns into quadratic equation

$$p_0^2 (1 - d_{il} x_{0,l} d_{ik} x_{0,k}) + 2p_0 d_{il} (p_{>} x_{>,l}) d_{ik} x_{0,k} - d_{il} (p_{>} x_{>,l}) d_{ik} (p_{>} x_{>,k}) - (-1)^p \gamma^2 \zeta(x) = 0.$$

It has two real roots of opposite signs iff $d_{il} x_{0,l} d_{ik} x_{0,k} < 1$. The sufficient condition is that the norm of $x_{i,k}$ as a linear operator is greater than the length of pD vector $x_{0,l}$. If $x_{0,l} = 0$ the simple answer exists:

$$E_p = E_p(x, p_{>}) = \sqrt{\sum_{i=1}^p (p_i(x, p_{>}))^2 + p_{>}^2 + (-1)^p \gamma^2 \zeta(x)}.$$

For the wave function, taking (6.288) into account, we get path integrals

$$\begin{aligned} \psi(x) &= \int \mathcal{D}^{n+1} p \mathcal{D}^{n+1} \tilde{x} \exp(i[p_\mu \Delta x^\mu - \\ &\quad - \omega u(p_0 + E_p(\tilde{x}, p_\perp)) - \omega v_i p_\mu x_i^\mu]) \psi(\tilde{x}) = \\ &= \int \mathcal{D}^{n-p} p_{>} \mathcal{D}^{n+1} \tilde{x} \exp(i[-p_{>} \Delta x_{>} - \omega u E_p(\tilde{x}, p_{>}) + \omega v_i p_{>} x_{>,i}]) \\ &\quad \times \delta(\Delta x^0 - \omega u - \omega v_i x'_{0,i}) \prod_{l=1}^r \delta(-\Delta x_l + \omega v_i x_{l,i}) \psi(\tilde{x}). \end{aligned}$$

Again δ -functions determine the *Lagrange multipliers*

$$\omega v_i = d_{il} \Delta x_l, \quad \text{and} \quad \omega u = \Delta x^0 - \omega v_i x_{0,i},$$

and reduce the number of integrals over x :

$$\begin{aligned} \psi(x) = & \int \mathcal{D}^{n-p} p_{>} \mathcal{D}^{n-p} \tilde{x}_{>} \exp(i[-p_{>} \Delta x_{>} \\ & - (\Delta x^0 - d_{il}(\Delta x_l) x_{0,i}) E_p(\tilde{x}, p_{>}) + d_{il}(\Delta x_l) p_{>} x_{>,i}]) \psi(\tilde{x}). \end{aligned}$$

Now we are ready to prove this statement without fixing any special coordinate system. We just need to find out the general formula for \vec{v}_\perp . Notice that by definition p^μ lies in a hyperplane of $x_{,0}^\mu$ and $x_{,i}^\mu$. Then, due to the constraints (6.297), p^μ is proportional to $x_{\perp,0}^\mu$. We have

$$v^\mu = \frac{\partial x^\mu(x^0, \sigma_i)}{\partial x^0} = \frac{x_{,0}^\mu}{x_{,0}^0},$$

and hence p^μ is also proportional to v_\perp^μ : $v_\perp^\mu = \alpha p^\mu$. To find the coefficient α we take

$$v_\perp^\mu p_\mu = \frac{x_{\perp,0}^\mu p_\mu}{x_{,0}^0} = \frac{-\gamma \sqrt{(-1)^p g}}{x_{,0}^0},$$

(the last equality is just the Euler's homogeneous function theorem) and

$$p^\mu p_\mu = (-1)^p \gamma^2 \zeta(x).$$

However [Golovnev and Prokhorov (2005)],

$$\text{as} \quad v_\perp^\mu p_\mu = \alpha p^\mu p_\mu, \quad \text{we have} \quad v_\perp^\mu = -\frac{\sqrt{(-1)^p g}}{(-1)^p \gamma \zeta x_{,0}^0} p^\mu,$$

$$\text{and} \quad 1 - \vec{v}_\perp^2 = v_\perp^\mu v_{\mu\perp} = \frac{(-1)^p g p^2}{\gamma^2 \zeta^2 x_{,0}^{0\,2}} = \frac{g}{\zeta x_{,0}^{0\,2}}, \quad \text{so we get}$$

$$\begin{aligned} \int dx^0 dV \sqrt{1 - \vec{v}_\perp^2} &= \int x_{,0}^0 d\sigma_0 \sqrt{|\det x_{,i}^\mu x_{\mu,k}|} \times \\ &\times d^p \sigma_i \sqrt{\frac{g}{\zeta x_{,0}^{0\,2}}} = \int d^{p+1} \sigma \sqrt{|g|}. \end{aligned}$$

6.7 Application: Topological String Theory

6.7.1 Quantum Geometry Framework

To start our review on topological string theory, here we depict a general *quantum geometry* framework (see e.g., [Witten (1998a)]).

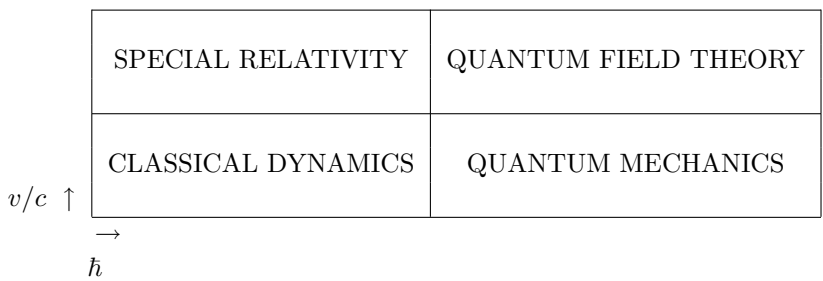


Fig. 6.22 The deformation from classical dynamics to quantum field theory (see text for explanation).

The relationship between non-relativistic classical mechanics and quantum field theory (see [Coleman (1988)]) can be summarized as in Figure 6.22. We see that the horizontal axis corresponds to the *Planck constant* \hbar (divided by the typical action of the system being studied), while the vertical axis corresponds to v/c , the ration of motion velocity and light velocity.

Similarly, in the superstring theory there are also two relevant expansion parameters, as shown in Figure 6.23. Here we see that the horizontal axis corresponds to the value of the *string coupling constant*, g_s , while the vertical axis corresponds to the value of the dimensionless *sigma model coupling* α'/R^2 with R being a typical radius of a compactified portion of space). In the extreme $\alpha' = g_s = 0$ limit, for instance, we recover relativistic particle dynamics. For nonzero g_s we recover point particle quantum field theory. For $g_s = 0$ and nonzero α' we are studying classical string theory. In general though, we need to understand the theory for arbitrary values of these parameters (see [Greene (1996)]).

Quantum stringy geometry postulates the existence of 6D *Calabi–Yau manifolds* at every point of the space–time (see, e.g., [Candelas *et. al.* (1985)]). These curled–up local manifolds transform according to the general *orbifolding* procedure, as will be described below.

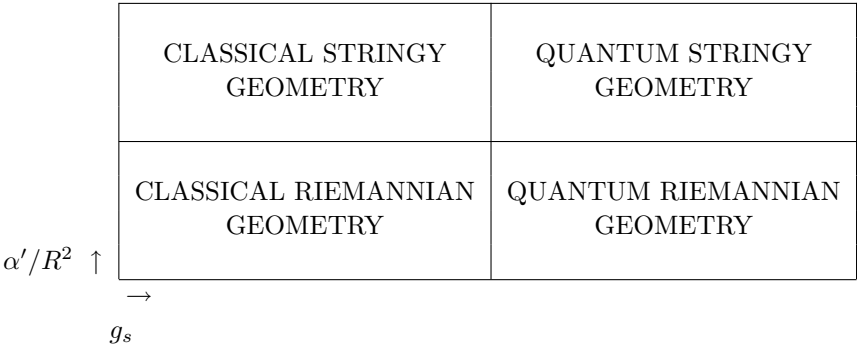


Fig. 6.23 The deformation from classical Riemannian geometry to *quantum stringy geometry* (see text for explanation).

6.7.2 Green–Schwarz Bosonic Strings and Branes

Here, we briefly describe the *world-sheet dynamics* of the *Green–Schwarz bosonic string theory*, and (more generally), bosonic p -brane theory, the predecessor of the current superstring theory (see [Schwarz (1993); Green *et. al.* (1987)] for details).

World–Line Description of a Point Particle

Recall that a point particle sweeps out a trajectory called world–line in space–time. This can be described by functions $x^\mu(\tau)$, that describe how the world–line, parameterized by τ , is embedded in the space–time, whose coordinates are denoted x^μ ($\mu = 0, 1, 2, 3$). For simplicity, let us assume that the space–time is flat Minkowski space with a *Lorentz metric tensor*

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Then, the Lorentz–invariant line element (metric form) is given by

$$ds^2 = -\eta_{\mu\nu} dx^\mu dx^\nu.$$

In normal units ($\hbar = c = 1$), the *action* for a particle of mass m is given by

$$S = -m \int ds.$$

This could be generalized to a curved space-time by replacing $\eta_{\mu\nu}$ by a Riemannian metric tensor $g_{\mu\nu}(x)$, but (for simplicity) we will not do so here. In terms of the embedding functions, $x^\mu(t)$, the action can be rewritten as

$$S[x] = -m \int d\tau \sqrt{-\eta_{\mu\nu} \dot{x}^\mu \dot{x}^\nu},$$

where overdot represents the derivative with respect to τ . An important property of this action is invariance under local reparametrizations. This is a kind of gauge invariance, whose meaning is that the form of S is unchanged under an arbitrary reparametrization of the world-line $\tau \rightarrow \tau(\tilde{\tau})$. Actually, one should require that the function $\tau(\tilde{\tau})$ is smooth and monotonic ($\frac{d\tau}{d\tilde{\tau}} > 0$). The reparametrization invariance is a 1D analog of the 4D general coordinate invariance of general relativity. Mathematicians refer to this kind of symmetry as diffeomorphism invariance.

The reparametrization invariance of S allows us to choose a gauge. A nice choice is the *static gauge*, $x^0 = \tau$. In this gauge (renaming the parameter to t) the action becomes

$$S = -m \int \sqrt{1 - v_i^2} dt, \quad \text{where} \quad v_i = \frac{dx_i}{dt}.$$

Requiring this action to be stationary under an arbitrary variation of $x_i(t)$ gives the Euler-Lagrangian equations

$$\frac{dp_i}{dt} = 0, \quad \text{where} \quad p_i = \frac{\delta S}{\delta v_i} = \frac{mv_i}{\sqrt{1 - v_i^2}},$$

which is the usual result. So we see that usual relativistic kinematics follows from the action $S = -m \int ds$.

***p*-Branes and World-Volume Actions**

We can now generalize the analysis of the massive point particle to a generic *p-brane*, which is characterized by its *tension* T_p . The action in this case involves the invariant $(p+1)$ D volume and is given by

$$S_p = -T_p \int d\mu_{p+1},$$

where the invariant volume element is

$$d\mu_{p+1} = \sqrt{-\det(-\eta_{\mu\nu} \partial_\alpha x^\mu \partial_\beta x^\nu)} d^{p+1}\sigma.$$

Here the embedding of the p -brane into dD space-time is given by functions $x^\mu(\sigma^\alpha)$. The index $\alpha = 0, \dots, p$ labels the $p+1$ coordinates σ^α of the p -brane world-volume and the index $\mu = 0, \dots, d-1$ labels the d coordinates x^μ of the dD space-time. We have defined $\partial_\alpha x^\mu = \frac{\partial x^\mu}{\partial \sigma^\alpha}$. The determinant operation acts on the $(p+1) \times (p+1)$ matrix whose rows and columns are labelled by α and β . The tension T_p is interpreted as the mass per unit volume of the p -brane. For a 0-brane, it is just the mass.

Let us now specialize to the string, $p = 1$. Evaluating the determinant gives the *Nambu-Goto action* (see subsection 6.5.4 above)

$$S[x] = -T \int d\sigma d\tau \sqrt{\dot{x}^2 x'^2 - (\dot{x} \cdot x')^2},$$

where we have defined $\sigma^0 = \tau$, $\sigma^1 = \sigma$, and $\dot{x}^\mu = \frac{\partial x^\mu}{\partial \tau}$, $x'^\mu = \frac{\partial x^\mu}{\partial \sigma}$. The above action is equivalent to the action

$$S[x, h] = -\frac{T}{2} \int d^2\sigma \sqrt{-h} h^{\alpha\beta} \eta_{\mu\nu} \partial_\alpha x^\mu \partial_\beta x^\nu, \quad (6.299)$$

where $h_{\alpha\beta}(\sigma, \tau)$ is the world-sheet metric, $h = \det h_{\alpha\beta}$, and $h^{\alpha\beta} = (h_{\alpha\beta})^{-1}$ is the inverse of $h_{\alpha\beta}$. The *Euler-Lagrangian equations* obtained by varying $h^{\alpha\beta}$ are

$$T_{\alpha\beta} = \partial_\alpha x \cdot \partial_\beta x - \frac{1}{2} h_{\alpha\beta} h^{\gamma\delta} \partial_\gamma x \cdot \partial_\delta x = 0.$$

In addition to reparametrization invariance, the action $S[x, h]$ has another local symmetry, called *conformal invariance*, or, *Weyl invariance*. Specifically, it is invariant under the replacement

$$h_{\alpha\beta} \rightarrow \Lambda(\sigma, \tau) h_{\alpha\beta}, \quad x^\mu \rightarrow x^\mu.$$

This local symmetry is special to the $p = 1$ case (strings).

The two reparametrization invariance symmetries of $S[x, h]$ allow us to choose a gauge in which the three functions $h_{\alpha\beta}$ (this is a symmetric 2×2 matrix) are expressed in terms of just one function. A convenient choice is the *conformally flat gauge*

$$h_{\alpha\beta} = \eta_{\alpha\beta} e^{\phi(\sigma, \tau)}.$$

Here, $\eta_{\alpha\beta}$ denoted the 2D Minkowski metric of a flat world-sheet. However, $h_{\alpha\beta}$ is only ‘conformally flat’, because of the factor e^ϕ . Classically,

substitution of this gauge choice into $S[x, h]$ leaves the gauge-fixed action

$$S = \frac{T}{2} \int d^2\sigma \eta^{\alpha\beta} \partial_\alpha x \cdot \partial_\beta x. \quad (6.300)$$

Quantum mechanically, the story is more subtle. Instead of eliminating h via its classical field equations, one should perform a Feynman path integral, using standard machinery to deal with the local symmetries and gauge fixing. When this is done correctly, one finds that in general ϕ does not decouple from the answer. Only for the special case $d = 26$ does the quantum analysis reproduce the formula we have given based on classical reasoning. Otherwise, there are correction terms whose presence can be traced to a conformal anomaly (i.e., a quantum-mechanical breakdown of the conformal invariance).

The gauge-fixed action is quadratic in the x 's. Mathematically, it is the same as a theory of d free scalar fields in two dimensions. The equations of motion obtained by varying x^μ are free 2D wave equations:

$$\ddot{x}^\mu - x''^\mu = 0.$$

However, this is not the whole story, because we must also take account of the constraints $T_{\alpha\beta} = 0$, which evaluated in the conformally flat gauge, read

$$T_{01} = T_{10} = \dot{x} \cdot x' = 0, \quad T_{00} = T_{11} = \frac{1}{2}(\dot{x}^2 + x'^2) = 0.$$

Adding and subtracting gives

$$(\dot{x} \pm x')^2 = 0. \quad (6.301)$$

Boundary Conditions

To go further, one needs to choose boundary conditions. There are three important types. For a closed string one should impose periodicity in the spatial parameter σ . Choosing its range to be π (as is conventional)

$$x^\mu(\sigma, \tau) = x^\mu(\sigma + \pi, \tau).$$

For an open string (which has two ends), each end can be required to satisfy either *Neumann* or *Dirichlet boundary conditions* for each value of μ ,

$$\begin{array}{ll} \text{Neumann :} & \frac{\partial x^\mu}{\partial \sigma} = 0 \quad \text{at } \sigma = 0 \text{ or } \pi, \\ \text{Dirichlet :} & \frac{\partial x^\mu}{\partial \tau} = 0 \quad \text{at } \sigma = 0 \text{ or } \pi. \end{array}$$

The Dirichlet condition can be integrated, and then it specifies a space–time location on which the string ends. The only way this makes sense is if the open string ends on a physical object – it ends on a D–brane.²⁶ If all the open–string boundary conditions are Neumann, then the ends of the string can be anywhere in the space–time. The modern interpretation is that this means that there are space–time–filling D–branes present.

Let us now consider the closed–string case in more detail. The general solution of the 2D wave equation is given by a sum of ‘right–movers’ and ‘left–movers’: $x^\mu(\sigma, \tau) = x_R^\mu(\tau - \sigma) + x_L^\mu(\tau + \sigma)$. These should be subject to the following additional conditions:

- $x^\mu(\sigma, \tau)$ is real,
- $x^\mu(\sigma + \pi, \tau) = x^\mu(\sigma, \tau)$, and
- $(x'_L)^2 = (x'_R)^2 = 0$ (these are the $T_{\alpha\beta} = 0$ constraints in (6.301)).

The first two of these conditions can be solved explicitly in terms of Fourier series:

$$x_R^\mu = \frac{1}{2}x^\mu + \ell_s^2 p^\mu(\tau - \sigma) + \frac{i}{\sqrt{2}}\ell_s \sum_{n \neq 0} \frac{1}{n} \alpha_n^\mu e^{-2in(\tau - \sigma)}$$

$$x_L^\mu = \frac{1}{2}x^\mu + \ell_s^2 p^\mu(\tau + \sigma) + \frac{i}{\sqrt{2}}\ell_s \sum_{n \neq 0} \frac{1}{n} \tilde{\alpha}_n^\mu e^{-2in(\tau + \sigma)},$$

where the expansion parameters α_n^μ , $\tilde{\alpha}_n^\mu$ satisfy $\alpha_{-n}^\mu = (\alpha_n^\mu)^\dagger$, $\tilde{\alpha}_{-n}^\mu = (\tilde{\alpha}_n^\mu)^\dagger$. The center–of–mass coordinate x^μ and momentum p^μ are also real. The fundamental string length scale ℓ_s is related to the tension T by

$$T = \frac{1}{2\pi\alpha'}, \quad \alpha' = \ell_s^2.$$

The parameter α' is called the universal Regge slope, since the string modes lie on linear parallel Regge trajectories with this slope.

Canonical Quantization

The analysis of closed–string left–moving modes, closed–string right–moving modes, and open–string modes are all very similar. Therefore, to avoid repetition, we focus on the closed–string right–movers. Starting with the gauge–fixed action in (6.300), the canonical momentum of the string is

$$p^\mu(\sigma, \tau) = \frac{\delta S}{\delta \dot{x}^\mu} = T \dot{x}^\mu.$$

²⁶D here stands for *Dirichlet*.

Canonical quantization (this is just free 2D field theory for scalar fields) gives

$$[p^\mu(\sigma, \tau), x^\nu(\sigma', \tau)] = -i\hbar\eta^{\mu\nu}\delta(\sigma - \sigma').$$

In terms of the Fourier modes (setting $\hbar = 1$) these become

$$[p^\mu, x^\nu] = -i\eta^{\mu\nu}, \quad [\alpha_m^\mu, \alpha_n^\nu] = m\delta_{m+n,0}\eta^{\mu\nu}, \quad [\tilde{\alpha}_m^\mu, \tilde{\alpha}_n^\nu] = m\delta_{m+n,0}\eta^{\mu\nu},$$

and all other commutators vanish.

Recall that a quantum-mechanical harmonic oscillator can be described in terms of raising and lowering operators, usually called a^\dagger and a , which satisfy $[a, a^\dagger] = 1$. We see that, aside from a normalization factor, the expansion coefficients α_{-m}^μ and α_m^μ are raising and lowering operators. There is just one problem. As $\eta^{00} = -1$, the time components are proportional to oscillators with the wrong sign ($[a, a^\dagger] = -1$). This is potentially very bad, because such oscillators create states of negative norm, which could lead to an inconsistent quantum theory (with negative probabilities, etc.). Fortunately, as we will explain, the $T_{\alpha\beta} = 0$ constraints eliminate the negative-norm states from the physical spectrum.

The classical constraint for the right-moving closed-string modes, $(x'_R)^2 = 0$, has Fourier components

$$L_m = \frac{T}{2} \int_0^\pi e^{-2im\sigma} (x'_R)^2 d\sigma = \frac{1}{2} \sum_{n=-\infty}^{\infty} \alpha_{m-n} \cdot \alpha_n,$$

which are called *Virasoro operators*. Since α_m^μ does not commute with α_{-m}^μ , L_0 needs to be normal-ordered:

$$L_0 = \frac{1}{2}\alpha_0^2 + \sum_{n=1}^{\infty} \alpha_{-n} \cdot \alpha_n.$$

Here $\alpha_0^\mu = \ell_s p^\mu / \sqrt{2}$, where p^μ is the momentum.

6.7.3 Calabi–Yau Manifolds, Orbifolds and Mirror Symmetry

Calabi–Yau Manifolds

Fundamental geometrical objects in string theory are the *Calabi–Yau manifolds* [Calabi (1957); Yau (1978)]. Recall (from subsection 3.14.2

above) that a Calabi–Yau manifold is a *compact Ricci-flat Kähler manifold* with a vanishing first Chern class. A Calabi–Yau manifold of complex dimension n is also called a Calabi–Yau n -fold, which is a manifold with an $SU(n)$ holonomy, i.e., it admits a global nowhere vanishing holomorphic $(n, 0)$ -form.

For example, in one complex dimension, the only examples are family of *tori*. Note that the Ricci-flat metric on the torus is actually a flat metric, so that the holonomy is the trivial group $SU(1)$. In particular, 1D *Calabi–Yau manifolds* are also called elliptic curves.

In two complex dimensions, the torus T^4 and the $K3$ surfaces²⁷ are the only examples. T^4 is sometimes excluded from the classification of being a Calabi–Yau, as its holonomy (again the trivial group) is a proper subgroup of $SU(2)$, instead of being isomorphic to $SU(2)$. On the other hand, the holonomy group of a $K3$ surface is the full $SU(2)$ group, so it may properly be called a Calabi–Yau in 2D.

In three complex dimensions, classification of the possible Calabi–Yau manifolds is an open problem. One example of a 3D Calabi–Yau is the quintic threefold in CP^4 .

In string theory, the term *compactification* refers to ‘curling up’ the extra dimensions (6 in the superstring theory), usually on Calabi–Yau spaces or on orbifolds. The mechanism behind this type of compactification is described by the Kaluza–Klein theory.

In the most conventional superstring models, 10 conjectural dimensions in string theory are supposed to come as 4 of which we are aware, carrying some kind of fibration with fiber dimension 6. Compactification on Calabi–Yau n -folds are important because they leave some of the original supersymmetry unbroken. More precisely, compactification on a Calabi–Yau 3-fold (with real dimension 6) leaves one quarter of the original supersymmetry unbroken.

Orbifolds

Recall that in topology, an *orbifold* is a generalization of a manifold, a topological space (called the underlying space) with an orbifold structure. The underlying space locally looks like a quotient of a Euclidean space under the action of a finite group of isometries.

²⁷Recall that $K3$ surfaces are compact, complex, simply-connected surfaces, with trivial canonical line bundle, named after three algebraic geometers, Kummer, Kähler and Kodaira. Otherwise, they are *hyperkähler manifolds* of real dimension 4 with $SU(2)$ holonomy.

The formal orbifold definition goes along the same lines as a definition of manifold, but instead of taking domains in \mathbb{R}^n as the target spaces of charts one should take domains of finite quotients of \mathbb{R}^n . A (topological) orbifold O , is a Hausdorff topological space X with a countable base, called the *underlying space*, with an *orbifold structure*, which is defined by orbifold atlas, given as follows.

An *orbifold chart* is an open subset $U \subset X$ together with open set $V \subset \mathbb{R}^n$ and a continuous map $\varphi : U \rightarrow V$ which satisfy the following property: there is a finite group Γ acting by linear transformations on V and a homeomorphism $\theta : U \rightarrow V/\Gamma$ such that $\varphi = \theta \circ \pi$, where π denotes the projection $V \rightarrow V/\Gamma$. A collection of orbifold charts, $\{\varphi_i = U_i \rightarrow V_i\}$, is called the *orbifold atlas* if it satisfies the following properties:

- (i) $\cup_i U_i = X$;
- (ii) if $\varphi_i(x) = \varphi_j(y)$ then there is a neighborhood $x \in V_x \subset V_i$ and $y \in V_y \subset V_j$ as well as a homeomorphism $\psi : V_x \rightarrow V_y$ such that $\varphi_i = \varphi_j \circ \psi$.

The orbifold atlas defines the orbifold structure completely and we regard two orbifold atlases of X to give the same orbifold structure if they can be combined to give a larger orbifold atlas. One can add differentiability conditions on the gluing map in the above definition and get a definition of *smooth* (C^∞) *orbifolds* in the same way as it was done for manifolds.

The main example of underlying space is a quotient space of a manifold under the action of a finite group of diffeomorphisms, in particular manifold with boundary carries natural orbifold structure, since it is \mathbb{Z}_2 -factor of its double. A factor space of a manifold along a smooth S^1 -action without fixed points carries an orbifold structure. The orbifold structure gives a natural stratification by open manifolds on its underlying space, where one strata corresponds to a set of singular points of the same type.

Note that one topological space can carry many different orbifold structures. For example, consider the orbifold O associated with a factor space of a 2-sphere S^2 along a rotation by π . It is homeomorphic to S^2 , but the natural orbifold structure is different. It is possible to adopt most of the characteristics of manifolds to orbifolds and these characteristics are usually different from the correspondent characteristics of the underlying space. In the above example, its orbifold fundamental group of O is \mathbb{Z}_2 and its *orbifold Euler characteristic* is 1.

Manifold orbifolding denotes an operation of wrapping, or folding in the case of mirrors, to superimpose all equivalent points on the original manifold – to get a new one.

In string theory, the word *orbifold* has a new flavor. In physics, the

notion of an orbifold usually describes an object that can be globally written as a coset M/G where M is a manifold (or a theory) and G is a group of its isometries (or symmetries). In string theory, these symmetries do not need to have a geometric interpretation. The so-called *orbifolding* is a general procedure of string theory to derive a new string theory from an old string theory in which the elements of the group G have been identified with the identity. Such a procedure reduces the number of string states because the states must be invariant under G , but it also increases the number of states because of the extra twisted sectors. The result is usually a new, perfectly smooth string theory.

Mirror Symmetry

The so-called *mirror symmetry* is a surprising relation that can exist between two Calabi–Yau manifolds. It happens, usually for two such 6D manifolds, that the shapes may look very different geometrically, but nevertheless they are equivalent if they are employed as hidden dimensions of a (super)string theory. More specifically, mirror symmetry relates two manifolds M and W whose *Hodge numbers* $h^{1,1} = \dim H^{1,1}$ and $h^{1,2} = \dim H^{1,2}$ are swapped; string theory compactified on these two manifolds leads to identical physical phenomena (see [Greene (2000)]).

Strominger showed in [Strominger (1990)] that mirror symmetry is a special example of the so-called T -duality: the Calabi–Yau manifold may be written as a fiber bundle whose fiber is a 3D torus $T^3 = S^1 \times S^1 \times S^1$. The simultaneous action of T -duality on all three dimensions of this torus is equivalent to mirror symmetry.

Mirror symmetry allowed the physicists to calculate many quantities that seemed virtually incalculable before, by invoking the ‘mirror’ description of a given physical situation, which can be often much easier. Mirror symmetry has also become a very powerful tool in mathematics, and although mathematicians have proved many rigorous theorems based on the physicists’ intuition, a full mathematical understanding of the phenomenon of mirror symmetry is still lacking.

6.7.4 More on Topological Field Theories

Unfortunately, there is no such thing as a crash course in string theory, but the necessary background can be found in the classic two-volume monographs [Green *et. al.* (1987)] and [Polchinski (1998)]. A good introduction

to conformal field theory is given in the lecture notes [Schellekens (1996)]. The basics of topological string theory were laid out in a series of beautiful papers by E. Witten in 1990s [Witten (1988b); Witten (1988a); Witten (1988d); Witten (1989); Witten (1990); Witten (1991); Witten (1992); Witten (1995a); Witten (1991)], and more or less completed in a seminal paper [Bershadsky *et. al.* (1994)]. Reviews about topological *string* theory, usually also contain a discussion of topological field theory. The first one of these is the review [Dijkgraaf *et. al.* (1991)] from the same period. However, if we want to dig even deeper, there is the 900–page book [Hori *et. al.* (2003)], which discusses topological string theory from the point of view of mirror symmetry.

In particular, there exists a mathematically rigorous, axiomatic definition of topological field theories due to [Atiyah (1988a)]. Instead of giving this definition, we will define topological field theory in a more physically–intuitive way, but as a result somewhat less rigorous way.

Recall (from subsection 6.5.1 above) that the output of a QFT is given by its observables: correlation functions of products of operators,

$$\langle \mathcal{O}_1(x_1) \cdots \mathcal{O}_n(x_n) \rangle_b. \quad (6.302)$$

Here, the $\mathcal{O}_i(x)$ are *physical* operators of the theory. What one calls ‘physical’ is part of the definition of the theory, but it is important to realize that in general not all combinations of fields are viewed as physical operators. For example, in a gauge theory, we usually require the observables to arise from gauge–invariant operators. That is, $\text{Tr } F$ would be one of the \mathcal{O}_i , but $\text{Tr } A$ or A itself would not.

The subscript b in the above formula serves as a reminder that the correlation function is usually calculated in a certain background. That is, the definition of the theory may involve a choice of a Riemannian manifold M on which the theory lives, it may involve choosing a metric on M , it may involve choosing certain coupling constants, and so on.

The definition of a topological field theory is now as follows. Suppose that we have a quantum field theory where the background choices involve a choice of manifold M and a choice of metric h on M . Then the theory is called a *topological field theory* if the observables (6.302) do *not* depend on the choice of metric h . Let us stress that it is part of the definition that h is a background field – in particular, we do *not* integrate over h in the path integral. One may wonder what happens if, once we have a topological field theory, we do make the metric h dynamical and integrate

over it. This is exactly what we will do once we start considering topological *string* theories.

Note that the word ‘topological’ in the definition may be somewhat of a misnomer [Vonk (2005)]. The reason is that the above definition does not strictly imply that the observables depend only on the topology of M – there may be other background choices hidden in b on which they depend as well. For example, in the case of a complex manifold M , correlation functions will in general not only depend on the topology of M and its metric, but also on our specific way of combining the 2D real coordinates on M into d complex ones. This choice, a *complex structure*, is part of the background of the quantum field theory, and correlation functions in a topological field theory will in general still depend on it.

If our quantum field theory has general coordinate invariance, as we will usually assume to be the case, then the above definition has an interesting consequence. The reason is that in such a case we can do an arbitrary general coordinate transformation, changing both the coordinates on M and its metric, under which the correlation functions should be invariant. Then, using the topological invariance, we can transform back the metric to its old value. The combined effect is that we have only changed the x_i in (6.302). That is, in a generally coordinate invariant topological field theory, the observables do not depend on the insertion points of the operators.

Chern Classes

Inspired by the identification of a connection with a gauge field, let us consider the analogue of the *non-Abelian field strength*,

$$F = dA - A \wedge A,$$

where $A \wedge A$ is a shorthand for $A_{Ji}^I A_{Kj}^J dx^i \wedge dx^j$ (note that $A \wedge A \neq 0$). A short calculation shows that on the overlap of two patches of M (or equivalently, under a gauge transformation), this quantity transforms as

$$F_{(b)} = \Lambda_{(ba)} F_{(a)} \Lambda_{(ba)}^{-1},$$

from which we see that F can be viewed as a section of a true vector bundle of Lie algebra valued 2-forms. In particular, we can take its trace and get a genuine 2-form:

$$c_1 = \frac{i}{2\pi} \text{Tr}(F),$$

where the prefactor is convention. It can be seen that this 2-form is closed:

$$d(\text{Tr}(F)) = \text{Tr}(dF) = -\text{Tr}(d(A \wedge A)) = -\text{Tr}(dA \wedge A - A \wedge dA) = 0.$$

Therefore, we can take its cohomology class, for which we would like to argue that it is a topological invariant. Note that this construction is independent of the choice of coordinates on M . Moreover, it is independent of gauge transformations. However, on a general vector bundle there may be connections which cannot be reached in this way from a given connection. Changing to such a connection is called a ‘large gauge transformation’, and from what we have said it is not clear a priori that the Chern classes do not depend on this choice of equivalence class of connections. However, with some work we can also prove this fact. The invariant $[c_1]$ is called the *first Chern class*. In fact, it might be better to call it a ‘relative topological invariant’: given a base manifold M of fixed topology, we can topologically distinguish vector bundles over it by calculating the above cohomology class.

By taking the trace of F , we lose a lot of information. There turns out to be a lot more topological information in F , and it can be extracted by considering the expression

$$c(F) = \det \left(1 + \frac{iF}{2\pi} \right),$$

where 1 is the identity matrix of the same size as the elements of the Lie algebra of G . Again, it can be checked that this expression is invariant under a change of coordinates for M and under a change of connection. Since the matrix components inside the determinant consist of the 0-form 1 and the 2-form F , expanding the determinant will lead to an expression consisting of forms of all even degrees. One writes this as

$$c(F) = c_0(F) + c_1(F) + c_2(F) + \dots$$

The sum terminates either at the highest degree encountered in expanding the determinant, or at the highest allowed even form on M . Note that $c_0(F) = 1$, and $[c_1(F)]$ is exactly the first Chern class we defined above. The cohomology class of c_n is called the *n th Chern class*.

As an almost trivial example, let us consider the case of a product bundle $M \times W$. In this case there is a global section $g(x)$ of the principal bundle P , and we can use this to construct a connection $A = -gdg^{-1}$, so

that [Vonk (2005)]

$$F = -dgdg^{-1} - gdg^{-1}gdg^{-1} = -dgdg^{-1} + dgdg^{-1} = 0.$$

Thus, for a trivial bundle with this connection, $c_0 = 1$ and $c_n = 0$ for all $n > 0$.

Chern–Simons Theory

The easiest way to construct a topological field theory is to construct a theory where both the *action* S (or, *quantum measure* e^{iS}) and the fields do not include the metric at all. Such topological field theories are called ‘Schwarz–type’ topological field theories. This may sound like a trivial solution to the problem, but nevertheless it can lead to quite interesting results. To see this, let us consider familiar example: *Chern–Simons gauge theory* on a 3D manifold M – now from a physical point of view.

Recall from subsection 5.11.8 that Chern–Simons theory is a gauge theory – that is, it is constructed from a vector bundle E over the base space M , with a structure group (gauge group) G and a connection (gauge field) A . The Lagrangian of Chern–Simons theory is then given by

$$L = \text{Tr}(A \wedge dA - \frac{2}{3}A \wedge A \wedge A).$$

It is a straightforward exercise to check how this Lagrangian changes under the gauge transformation

$$\tilde{A} = gAg^{-1} - gdg^{-1},$$

and one finds

$$\begin{aligned} \tilde{L} &\equiv \text{Tr}(\tilde{A} \wedge d\tilde{A} - \frac{2}{3}\tilde{A} \wedge \tilde{A} \wedge \tilde{A}) \\ &= \text{Tr}(A \wedge dA - \frac{2}{3}A \wedge A \wedge A) - d \text{Tr}(gA \wedge dg^{-1}) + \frac{1}{3} \text{Tr}(gdg^{-1} \wedge dg \wedge dg^{-1}). \end{aligned}$$

The second term is a total derivative, so if M does not have a boundary, the action, being the integral of L over M , does not get a contribution from this term. The last term is not a total derivative, but its integral turns out to be a topological invariant of the map $g(x)$, which is quantized as

$$\frac{1}{24\pi^2} \int_M \text{Tr}(gdg^{-1} \wedge dg \wedge dg^{-1}) = m \in \mathbb{Z}.$$

From this, we see that if we define the *action* as

$$S = \frac{k}{4\pi} \int_M L,$$

with k an integer, the action changes by $2\pi km$ under gauge transformations, and the quantum measure e^{iS} is invariant.

So from this discussion we seem to arrive at the conclusion that the *partition function*

$$Z = \int \mathcal{D}[A] e^{iS[A]}$$

for a line bundle of a fixed topology E is a topological invariant of M , as are the correlation functions of gauge-invariant operators such as $\text{Tr } F$. However, there is one more detail we have to worry about: there may be an anomaly in the quantum theory. That is, it may not be possible to define the path integral measure $\mathcal{D}[A]$ in a gauge-invariant way.

One way to see what problems can arise is to note that to actually calculate the path integral, one has to pick a gauge condition on A . That is, we have to pick one representative of A in each equivalence class under gauge transformations. To make such a choice will in general require a choice of metric. For example, from electromagnetism (where E is a 1D complex line bundle and $G = U(1)$) we know that a useful gauge is the Feynman gauge, in which the equation of motion for A becomes

$$\Delta A = 0.$$

As we have seen before, the Laplacian Δ is an operator which, through the Hodge star, depends on the metric, and hence the results we find will a priori be metric dependent. To show that the results are truly metric independent, one needs to show that the quantum results do not depend on our arbitrary choice of gauge.

We will not go into the details of this, but state that one can show that Chern–Simons theory on a compact 3-manifold is anomaly-free, so our naive argument above was correct, and one can indeed calculate topological invariants of M in this way.

Let us briefly discuss the kind of topological invariants that Chern–Simons theory can lead to. Recall that one can construct a Lie group element g from a Lie algebra element \mathcal{A} as $g = e^{\mathcal{A}}$. Now suppose we have a path $\gamma(t)$ inside M . Suppose that we chop up γ into very small line elements given by tangent vectors $\dot{\gamma}\delta t$. Then we can insert this tangent vector into

the connection 1-form A , and get a Lie algebra element. As we have seen, it is precisely this Lie algebra element which transports vectors in E along this small distance: we have to multiply these vectors by $1 + \mathcal{A}$. This is a linear approximation to the finite transformation $e^{\mathcal{A}}$. So if we transport a vector along the entire closed curve γ , it will return multiplied by a group element

$$g = \lim_{\delta t \rightarrow 0} [\exp(A(\dot{\gamma}(0))\delta t) \exp(A(\dot{\gamma}(\delta t))\delta t) \exp(A(\dot{\gamma}(2\delta t))\delta t) \cdots].$$

Now it is tempting to add all the exponents and write their sum in the limit as an integral, but this is not quite allowed since the different group elements may not commute, so $e^X e^Y \neq e^{X+Y}$. Therefore, one uses the following notation,

$$g = P \exp \int_{\gamma} A,$$

where P stands for path ordering, while the element g is called the *holonomy* of A around the closed curve γ . An interesting gauge and metric independent object turns out to be the trace of this group element. This trace is called the *Wilson loop* $W_{\gamma}(A)$, given by

$$W_{\gamma}(A) = \text{Tr}(P) \exp \int_{\gamma} A.$$

The topological invariants we are interested in are now the correlation functions of such Wilson loops in Chern–Simons theory. Since these correlation functions are independent of the parametrization of M , we can equivalently say that they will be independent of the precise location of the loop γ ; we have in fact constructed a topological invariant of the embedding of γ inside M . This embedding takes the shape of a knot, so the invariants we have constructed are knot invariants. One can show that the invariants are actually polynomials in the variable $y = \exp 2\pi i/(k+2)$, where k is the integer ‘coupling constant’ of the Chern–Simons theory.

The above construction is due to E. Witten, and was carried out in [Witten (1988a)]. Before Witten’s work, several polynomial invariants of knots were known, one of the simplest ones being the so-called *Jones polynomial*. It can be shown that many of these polynomials arise as special cases of the above construction, where one takes a certain structure group G , $SU(2)$ for the Jones polynomial, and a certain vector bundle (representation) E , the fundamental representation for the Jones polynomial. That is, using this ‘trivial’ topological field theory, Witten was able to reproduce

a large number of the known knot invariants in a unified framework, and construct a great number of new invariants as well [Vonk (2005)].

Cohomological Field Theories

Even though the above example leads to quite interesting topological invariants, the construction itself is somewhat trivial: given the absence of anomalies that we mentioned, the independence of the metric is completely manifest throughout the procedure. There exists a different way of constructing topological field theories in which the definition of the theory *does* use a metric, but one can still show that the partition function and the physical correlation functions of the theory are metric-independent. The theories constructed in this way are called topological theories of ‘Witten-type’, or cohomological field theories.

Cohomological field theories are field theories that possess a very special type of symmetry. Recall that from Noether’s Theorem a global symmetry of a theory leads to a conserved charge S , and that after quantizing the theory, the symmetry is generated by the corresponding operator:

$$\delta_\epsilon \mathcal{O}_i = i\epsilon[S, \mathcal{O}_i], \quad \text{or} \quad \delta_\epsilon \mathcal{O}_i = i\epsilon\{S, \mathcal{O}_i\},$$

depending on whether S and \mathcal{O}_i are fermionic or bosonic. Furthermore, the symmetry-invariant states $|j\rangle$ satisfy

$$S|j\rangle = 0. \tag{6.303}$$

In particular, if the symmetry is not spontaneously broken, the vacuum of such a theory will be symmetric, i.e., $S|0\rangle = 0$, and expectation values of operators will be unchanged after a symmetry transformation:

$$\langle 0|\mathcal{O}_i + \delta\mathcal{O}_i|0\rangle = \langle 0|\mathcal{O}_i|0\rangle + i\epsilon\langle 0|S\mathcal{O}_i \pm \mathcal{O}_i S|0\rangle = \langle 0|\mathcal{O}_i|0\rangle,$$

since S annihilates the vacuum. Note that, to linear order in a ‘small parameter’ ϵ , the second term in the first line (with $|0\rangle$ replaced by an arbitrary state $|\psi\rangle$) can also be obtained if instead of on the operators, we let the symmetry operator S act on the *state* as

$$|\psi\rangle \rightarrow |\psi\rangle + i\epsilon S|\psi\rangle. \tag{6.304}$$

In case S is the Hamiltonian, this is the infinitesimal version of the well-known transition between the Schrödinger and Heisenberg pictures. Note

that the equations (6.303) and (6.304) already contain some flavor of cohomology. Cohomological field theories are theories where this analogy can be made exact.

With this in mind, the first property of a cohomological field theory will not come as a surprise: it should contain a fermionic symmetry operator Q which squares to zero, $Q^2 = 0$. This may seem like a strange requirement for a field theory, but symmetries of this type occur for example when we have a gauge symmetry and fix it by using the Faddeev–Popov procedure; the resulting theory will then have a global *BRST symmetry*, which satisfies precisely this constraint. Another example is found in supersymmetry, where one also encounters symmetry operators that square to zero, as we will see in detail later on.

The second property a cohomological field theory should have is really a definition: we define the physical operators in this theory to be the operators that are closed under the action of this Q –operator²⁸:

$$\{Q, \mathcal{O}_i\} = 0. \quad (6.305)$$

Again, this may seem to be a strange requirement for a physical theory, but again it naturally appears in BRST quantization, and for example in conformal field theories, where we have a 1–1 correspondence between operators and states. In such theories, the symmetry requirement (6.303) on the states translates into the requirement (6.305) on the operators.

Thirdly, we want to have a theory in which the Q –symmetry is not spontaneously broken, so the vacuum is symmetric. Note that this implies the equivalence

$$\mathcal{O}_i \sim \mathcal{O}_i + \{Q, \Lambda\}. \quad (6.306)$$

The reason for this is that the expectation value of an operator product involving a Q –exact operator $\{Q, \Lambda\}$ takes the form

$$\langle 0 | \mathcal{O}_{i_1} \cdots \mathcal{O}_{i_j} \{Q, \Lambda\} \mathcal{O}_{i_{j+1}} \cdots \mathcal{O}_{i_n} | 0 \rangle = \langle 0 | \mathcal{O}_{i_1} \cdots \mathcal{O}_{i_j} (Q\Lambda - \Lambda Q) \mathcal{O}_{i_{j+1}} \cdots \mathcal{O}_{i_n} | 0 \rangle,$$

and each term vanishes separately, e.g.,

$$\begin{aligned} \langle 0 | \mathcal{O}_{i_1} \cdots \mathcal{O}_{i_j} Q\Lambda \mathcal{O}_{i_{j+1}} \cdots \mathcal{O}_{i_n} | 0 \rangle &= \pm \langle 0 | \mathcal{O}_{i_1} \cdots Q\mathcal{O}_{i_j} \Lambda \mathcal{O}_{i_{j+1}} \cdots \mathcal{O}_{i_n} | 0 \rangle \\ &= \pm \langle 0 | Q\mathcal{O}_{i_1} \cdots \mathcal{O}_{i_j} \Lambda \mathcal{O}_{i_{j+1}} \cdots \mathcal{O}_{i_n} | 0 \rangle = 0, \end{aligned} \quad (6.307)$$

²⁸From now on, we denote both the commutator and the anti-commutator by curly brackets, unless it is clear that one of the operators inside the brackets is bosonic.

where we made repeated use of (6.305). Together, (6.305) and (6.306) mean that our physical operators are Q -cohomology classes.

The fourth and final requirement for a cohomological field theory is that the *metric SEM-tensor* is Q -exact,

$$T_{\alpha\beta} \equiv \frac{\delta S}{\delta h^{\alpha\beta}} = \{Q, G_{\alpha\beta}\} \quad (6.308)$$

for some operator $G_{\alpha\beta}$. The physical interpretation of this is the following. The integrals of the components $T_{0\alpha}$ over a space-like hypersurface are conserved quantities. For example, the integral of T_{00} gives the Hamiltonian:

$$H = \int_{space} T_{00}.$$

Similarly, T_{0a} for $a \neq 0$ give the momentum charges. Certainly, the Hamiltonian H should commute with all symmetry operators of the theory, and one usually takes the other space-time symmetries to commute with the internal symmetries as well. The choice of the first lower index 0 here is related to a choice of Lorentz frame, so in general the integrals of all $T_{\alpha\beta}$ will commute with the internal symmetries. In a local theory, it is then natural to assume that also the *densities* commute with Q . However, (6.308) is an even stronger requirement: the SEM densities should not only commute with Q (that is, be Q -closed), but they have to do so in a trivial way (they should be Q -commutators, that is, Q -exact) for the theory to be cohomological.

This fourth requirement is the crucial one in showing the topological invariance of the theory. Let us consider the functional $h^{\alpha\beta}$ -derivative of an observable:

$$\begin{aligned} \frac{\delta}{\delta h^{\alpha\beta}} \langle \mathcal{O}_{i_1} \cdots \mathcal{O}_{i_n} \rangle &= \frac{\delta}{\delta h^{\alpha\beta}} \left(\int D\phi \mathcal{O}_{i_1} \cdots \mathcal{O}_{i_n} e^{iS[\phi]} \right) \\ &= i \int D\phi \mathcal{O}_{i_1} \cdots \mathcal{O}_{i_n} \frac{\delta S}{\delta h^{\alpha\beta}} e^{iS[\phi]} = i \langle \mathcal{O}_{i_1} \cdots \mathcal{O}_{i_n} \{Q, G_{\alpha\beta}\} \rangle = 0, \end{aligned}$$

where in the last line we used the same argument as in (6.307). One might be worried about the operator ordering in going from the operator formalism to the path integral formalism and back. We should really have inserted a time ordering operator on the r.h.s. in the first step above. However, the result then shows us that we can arbitrarily change the metric – and hence the time ordering of the operators, so with hindsight we may actually think of the operators as being arbitrarily ordered. Finally, we have assumed that our operators do not depend explicitly on the metric.

A very practical way to ensure (6.308) is to use a Lagrangian which itself is Q -exact, $L = \{Q, V\}$, for some operator V . This choice has an extra virtue, which we can see if we explicitly include Planck's constant in our description: the quantum measure then reads

$$\exp \frac{i}{\hbar} \left\{ Q, \int_M V \right\}.$$

Then, we can use exactly the same argument as before to show that

$$\frac{d}{d\hbar} \langle \mathcal{O}_{i_1} \cdots \mathcal{O}_{i_n} \rangle = 0.$$

That is, the correlators we are interested in are independent of \hbar , and we can therefore calculate them *exactly* in the classical limit.

Descent equations

An important property of cohomological field theories is that, given a scalar physical operator on M – where by ‘scalar’ we mean an operator that does not transform under coordinate transformations of M , so in particular it has no α -indices – we can construct further operators which behave like p -forms on M . The basic observation is that we can integrate (6.308) over a spatial hypersurface to get a similar relation for the momentum operators:

$$P_\alpha = \{Q, G_\alpha\},$$

where G_α is a fermionic operator. Now consider the operator

$$\mathcal{O}_\alpha^{(1)} = i\{G_\alpha, \mathcal{O}^{(0)}\},$$

where $\mathcal{O}^{(0)}(x)$ is a scalar physical operator: $\{Q, \mathcal{O}^{(0)}(x)\} = 0$. Let us calculate

$$\begin{aligned} \frac{d}{dx^\alpha} \mathcal{O}^{(0)} &= i[P_\alpha, \mathcal{O}^{(0)}] = [\{Q, G_\alpha\}, \mathcal{O}^{(0)}] \\ &= \pm i\{\{G_\alpha, \mathcal{O}^{(0)}\}, Q\} - i\{\{\mathcal{O}^{(0)}, Q\}, G_\alpha\} = \{Q, \mathcal{O}_\alpha^{(1)}\}. \end{aligned}$$

In going from the second to the third line, we have used the Jacobi identity. The first sign in the third line depends on whether $\mathcal{O}^{(0)}$ is bosonic or fermionic, but there is no sign ambiguity in the last line. By defining the 1-form operator

$$\mathcal{O}^{(1)} = \mathcal{O}_\alpha^{(1)} dx^\alpha$$

we can write this result as

$$d\mathcal{O}^{(0)} = \{Q, \mathcal{O}^{(1)}\}.$$

Then, we can integrate this equation over a closed curve $\gamma \subset M$ to find

$$\{Q, \int_{\gamma} \mathcal{O}^{(1)}\} = 0.$$

That is, by constructing a $\int_{\gamma} \mathcal{O}^{(1)}$ for each $\mathcal{O}^{(0)}$, we have found a whole class of new, non-local, physical operators.

The above procedure can now be repeated in exactly the same way starting from $\mathcal{O}^{(1)}$, and doing this we find a whole tower of p -form operators:

$$\begin{aligned} \{Q, \mathcal{O}^{(0)}\} &= 0, & \{Q, \mathcal{O}^{(1)}\} &= d\mathcal{O}^{(0)}, & \{Q, \mathcal{O}^{(2)}\} &= d\mathcal{O}^{(1)}, \\ \cdots, & & \{Q, \mathcal{O}^{(n)}\} &= d\mathcal{O}^{(n-1)}, & 0 &= d\mathcal{O}^{(n)}. \end{aligned}$$

The last equation is trivial, since there are no $(n+1)$ -forms on an nD smooth manifold.

Following the same reasoning, the integral of $\mathcal{O}^{(p)}$ over a pD submanifold of M is now a physical operator. This gives us a large class of new physical operators, starting from the scalar ones. Note that these operators, being integrated over a submanifold, are inherently nonlocal. Nevertheless, they can have a very physical interpretation. Particularly important examples of this are the ‘top-form’ operators $\mathcal{O}^{(n)}$ that can be integrated over the whole manifold, leading to

$$\left\{Q, \int_M \mathcal{O}^{(n)}\right\} = 0.$$

This implies that we can add terms $t^a \mathcal{O}_a^{(n)}$, with t^a arbitrary coupling constants, to our Lagrangian without spoiling the fact that the theory is cohomological. These deformations of the theory will be important to us later.

2D Cohomological Field Theories

Since string theories are 2D field theories, we will in particular be interested in cohomological field theories in two dimensions (see, e.g., [Witten (1995b)]). These theories have some extra properties which will be important in our discussion. Let us begin by reminding the reader of the relation between states in the operator formalism of quantum field theories, and

boundary conditions in the path–integral formalism. In its simplest form, this relation looks like

$$\int_{BC1}^{BC2} D\phi \cdots e^{iS[\phi]} = \langle BC1 | T(\cdots) | BC2 \rangle. \quad (6.309)$$

Here, we included a time–ordering operator for completeness, but as we have stated before, for the theories we are interested in this ordering is irrelevant. The notation $|BCi\rangle$ indicates the state corresponding to the incoming or outgoing boundary condition. For example, if on the path integral side we prescribe all fields at a certain initial time, $\phi(t = t_1) = f(t_1)$, on the operator side this corresponds to the incoming state satisfying

$$\phi(t_1)|BC1\rangle = f(t_1)|BC1\rangle. \quad (6.310)$$

where on the l.h.s. we have an operator acting, but on the r.h.s. there is a simple scalar multiplication. More generally, in the operator formalism we can have linear combinations of states of the type (6.310). Therefore, we should allow for linear combinations on the l.h.s. of (6.309) as well. In other words, in the path–integral formalism, *a state is an operator which adds a number (a weight) to each possible boundary condition on the fields*. From this point of view, the states in (6.310) are like ‘delta–functionals’: they assign weight 1 to the boundary condition $\phi(t_1) = f(t_1)$, and weight 0 to all other boundary conditions.

Now, let us specialize to 2D field theories. Here, the boundary of a compact manifold is a set of circles. Let us for simplicity assume that the ‘incoming’ boundary consists of a single circle. We can now define a state in the above sense by doing a path integral over a second surface with the topology of a hemisphere. This path integral gives a number for every boundary condition of the fields on the circle, and this is exactly what a state in the path–integral formalism should do. In particular, one can use this procedure to define a state corresponding to every operator \mathcal{O}_a by inserting \mathcal{O}_a on the hemisphere and then stretching this hemisphere to infinite size. An expectation value in the operator formalism, such as

$$\langle \mathcal{O}_a | \mathcal{O}_b(x_2) \mathcal{O}_c(x_3) | \mathcal{O}_d \rangle_{cyl}, \quad (6.311)$$

on a cylinder of finite length, can then schematically be drawn as in Figure 6.24. Here, instead of first doing the path integrals over the semi–infinite hemispheres and inserting the result in the path integral over the cylinder, one can just as well integrate over the whole surface at once. However, note that in topological field theories, there is no need to do the stretching, since

the path integral only depends on the topology of the surface, and hence not on the size of the hemisphere [Vonk (2005)].

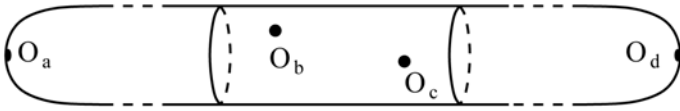


Fig. 6.24 A graphical representation of the correlation function (6.311).

We assume that all states that we are interested in are of the above form, which in particular means that we will integrate only over Riemann surfaces without boundary. Moreover, we assume these surfaces to be orientable. An important property of topological field theories in two dimensions is now that its correlation functions factorize in the following way:

$$\langle \mathcal{O}_1 \cdots \mathcal{O}_n \rangle_\Sigma = \sum_{a,b} \langle \mathcal{O}_1 \cdots \mathcal{O}_i \mathcal{O}_a \rangle_{\Sigma_1} \eta^{ab} \langle \mathcal{O}_b \mathcal{O}_{i+1} \cdots \mathcal{O}_n \rangle_{\Sigma_2}, \quad (6.312)$$

where the genus of Σ is the sum of the genera of Σ_1 and Σ_2 . This statement is explained in Figure 6.25. By using the topological invariance, we can deform a Riemann surface Σ with a set of operator insertions in such a way that it develops a long tube. From general quantum field theory, we know that if we stretch this tube long enough, only the asymptotic states – that is, the states in the physical part of the Hilbert space – will propagate. But as we have just argued, instead of inserting these asymptotic states, we may just as well insert the corresponding operators at a finite distance. However, to conclude that this leads to (6.312), we have to show that this definition of ‘physical states’ – being the ones that need to be inserted as asymptotic states – agrees with our previous definition in terms of Q -cohomology. Let us argue that it does by first writing the factorization as

$$\langle \mathcal{O}_1 \cdots \mathcal{O}_n \rangle_\Sigma = \sum_{A,B} \langle \mathcal{O}_1 \cdots \mathcal{O}_i | \mathcal{O}_A \rangle_{\Sigma_1} \eta^{AB} \langle \mathcal{O}_B | \mathcal{O}_{i+1} \cdots \mathcal{O}_n \rangle_{\Sigma_2}. \quad (6.313)$$

where the \mathcal{O}_A with capital index now correspond to a *complete* basis of asymptotic states in the Hilbert space. The reader may be more used to this type of expressions in the case where $\eta_{AB} = \delta_{AB}$, but since we have not shown that with our definitions $\langle \mathcal{O}_A | \mathcal{O}_B \rangle = \delta_{AB}$, we have to work with this more general form of the identity operator, where η is a metric that we will determine in a moment.

Now, we can write the Hilbert space as a direct sum, $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$, where \mathcal{H}_0 consists of states $|\psi\rangle$ for which $Q|\psi\rangle = 0$ and \mathcal{H}_1 is its orthogonal complement. Since

$$Q(\mathcal{O}_1 \dots \mathcal{O}_i | 0\rangle) = 0, \quad (6.314)$$

the states in \mathcal{H}_1 are in particular orthogonal to states of the form $\mathcal{O}_1 \dots \mathcal{O}_i | 0\rangle$, and hence the states in \mathcal{H}_1 do not contribute to the sum in (6.313). Moreover, changing \mathcal{O}_A to $\mathcal{O}_A + \{Q, \Lambda\}$ does not change the result in (6.313), so we only need to sum over a basis of $\mathcal{H}_0/\mathfrak{Z}(\{Q, \cdot\})$, which is exactly the space \mathcal{H}_{phys} of ‘topologically physical’ states.

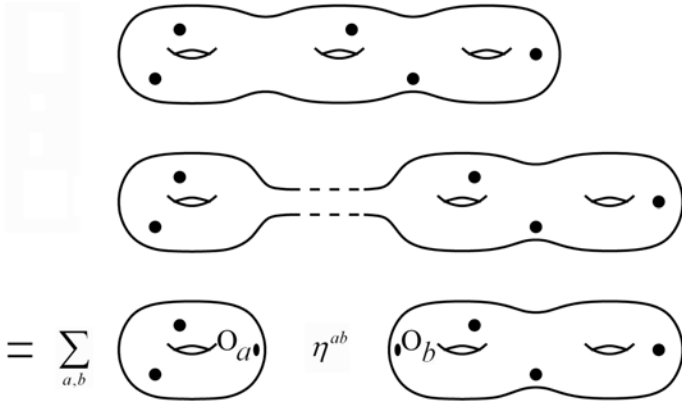


Fig. 6.25 A correlation function on a Riemann surface factorizes into correlation functions on two Riemann surfaces of lower genus (see text for explanation).

Finally, let us determine the metric η^{ab} . We can deduce its form by factorizing the 2-point function

$$C_{ab} = \langle \mathcal{O}_a \mathcal{O}_b \rangle \quad (6.315)$$

in the above way, resulting in

$$C_{ab} = C_{ac} \eta^{cd} C_{db}. \quad (6.316)$$

In other words, we find that the metric η^{ab} is the matrix inverse of the 2-point function C_{ab} , which for this reason we will write as η_{ab} from now on.

One can apply a similar procedure to ‘cut open’ internal loops in a Riemann surface Σ , so that we get

$$\langle \mathcal{O}_1 \cdots \mathcal{O}_n \rangle_\Sigma = (-1)^{F_a} \eta^{ab} \sum_{a,b} \langle \mathcal{O}_a \mathcal{O}_b \mathcal{O}_1 \cdots \mathcal{O}_n \rangle_{\Sigma'}, \quad (6.317)$$

where the genus of Σ' is one less than the genus of Σ . The factor $(-1)^{F_a}$ multiplies the expression on the r.h.s. by -1 if the inserted operator \mathcal{O}_a (and hence also \mathcal{O}_b) is fermionic. Proving that it needs to be included is not straightforward, but one can think of it as the ‘stringy version’ of the well-known statement from quantum field theory that fermion loops add an extra minus sign to a Feynman diagram.

The reader should convince himself that together, the equations (6.312) and (6.317) imply that we can reduce all n -point correlation functions to products of 3-point functions on the sphere. We denote these important quantities by

$$c_{abc} \equiv \langle \mathcal{O}_a \mathcal{O}_b \mathcal{O}_c \rangle_0, \quad (6.318)$$

where the label 0 denotes the genus of the sphere. By using (6.312) to separate two insertion points on a sphere, we see that

$$\langle \cdots \mathcal{O}_a \mathcal{O}_b \cdots \rangle_\Sigma = \sum_{c,d} \langle \cdots \mathcal{O}_c \cdots \rangle_\Sigma \eta^{cd} c_{abd} = \sum_c \langle \cdots \mathcal{O}_c \cdots \rangle_\Sigma c_{ab}{}^c,$$

where we raised an index of c_{abc} with the metric η . We can view the above result as the definition of an operator algebra with structure constants $c_{ab}{}^c$:

$$\mathcal{O}_a \mathcal{O}_b = \sum_c c_{ab}{}^c \mathcal{O}_c. \quad (6.319)$$

From the metric η_{ab} and the structure constants $c_{ab}{}^c$, we can now calculate any correlation function in the cohomological field theory.

6.7.5 Topological Strings

The 2D field theories we have constructed are already very similar to string theories. However, one ingredient from string theory is missing: in string theory, the world-sheet theory does not only involve a path integral over the maps ϕ^i to the target space and their fermionic partners, but also a path integral over the world-sheet metric $h_{\alpha\beta}$. So far, we have set this metric to a fixed background value.

We have also encountered a drawback of our construction. Even though the theories we have found can give us some interesting ‘semi-topological’ information about the target spaces, one would like to be able to define general nonzero n -point functions at genus g instead of just the partition function at genus one and the particular correlation functions we calculated at genus zero.

It turns out that these two remarks are intimately related. In this section we will go from topological field theory to topological string theory by introducing integrals over all metrics, and in doing so we will find interesting nonzero correlation functions at any genus (see [Vonk (2005)]).

Coupling to Topological Gravity

In coupling an ordinary field theory to gravity, one has to perform the following three steps.

- First of all, one rewrites the Lagrangian of the theory in a covariant way by replacing all the flat metrics by the dynamical ones, introducing covariant derivatives and multiplying the measure by a factor of $\sqrt{\det h}$.
- Secondly, one introduces an Einstein–Hilbert term as the ‘kinetic’ term for the metric field, plus possibly extra terms and fields to preserve the symmetries of the original Lagrangian.
- Finally, one has to integrate the resulting theory over the space of all metrics.

Here we will not discuss the first two steps in this procedure. As we have seen in our discussion of topological field theories, the precise form of the Lagrangian only plays a comparatively minor role in determining the properties of the theory, and we can derive many results without actually considering a Lagrangian. Therefore, let us just state that it is possible to carry out the analog of the first two steps mentioned above, and construct a Lagrangian with a ‘dynamical’ metric which still possesses the topological Q -symmetry we have constructed. The reader who is interested in the details of this construction is referred to the paper [Witten (1990)] and to the lecture notes [Dijkgraaf *et. al.* (1991)].

The third step, integrating over the space of all metrics, is the one we will be most interested in here. Naively, by the metric independence of our theories, integrating their partition functions over the space of all metrics, and then dividing the results by the volume of the topological ‘gauge group’,

would be equivalent to multiplication by a factor of 1,

$$Z[h_0] \stackrel{?}{=} \frac{1}{G_{top}} \int \mathcal{D}[h] Z[h], \quad (6.320)$$

for any arbitrary background metric h_0 . There are several reasons why this naive reasoning might go wrong:

- There may be metric configurations which cannot be reached from a given metric by continuous changes.
- There may be anomalies in the topological symmetry at the quantum level preventing the conclusion that all gauge fixed configurations are equivalent.
- The volume of G_{top} is infinite, so even if we could rigorously define a path integral the above multiplication and division would not be mathematically well-defined.

For these reasons, we should really be more careful and precisely define what we mean by the ‘*integral over the space of all metrics*’. Let us note the important fact that just like in ordinary string theory (and even before twisting), the 2D sigma models become conformal field theories when we include the metric in the Lagrangian. This means that we can borrow the technology from string theory to integrate over all conformally equivalent metrics. As is well known, and as we will discuss in more detail later, the conformal symmetry group is a huge group, and integrating over conformally equivalent metrics leaves only a nD integral over a set of world-sheet moduli. Therefore, our strategy will be to use the analogy to ordinary string theory to first do this integral over all conformally equivalent metrics, and then perform the integral over the remaining nD moduli space.

In integrating over conformally equivalent metrics, one usually has to worry about conformal anomalies. However, here a very important fact becomes our help. To understand this fact, it is useful to rewrite our twisting procedure in a somewhat different language (see [Vonk (2005)]).

Let us consider the SEM-tensor $T_{\alpha\beta}$, which is the conserved Noether current with respect to global translations on \mathbb{C} . From conformal field theory, it is known that $T_{z\bar{z}} = T_{\bar{z}z} = 0$, and the fact that T is a conserved current, $\partial_\alpha T^\alpha_\beta = 0$, means that $T_{zz} \equiv T(z)$ and $T_{\bar{z}\bar{z}} \equiv \bar{T}(\bar{z})$ are (anti-)holomorphic in z . One can now expand $T(z)$ in Laurent modes,

$$T(z) = \sum L_m z^{-m-2}. \quad (6.321)$$

The L_m are called the *Virasoro generators*, and it is a well-known result from conformal field theory that in the quantum theory their commutation relations are

$$[L_m, L_n] = (m - n)L_{m+n} + \frac{c}{12}m(m^2 - 1)\delta_{m+n}.$$

The number c depends on the details of the theory under consideration, and it is called the central charge. When this central charge is nonzero, one runs into a technical problem. The reason for this is that the equation of motion for the metric field reads

$$\frac{\delta S}{\delta h^{\alpha\beta}} = T_{\alpha\beta} = 0.$$

In conformal field theory, one imposes this equation as a constraint in the quantum theory. That is, one requires that for physical states $|\psi\rangle$,

$$L_m|\psi\rangle = 0 \quad (\text{for all } m \in \mathbb{Z}).$$

However, this is clearly incompatible with the above commutation relation unless $c = 0$. In string theory, this value for c can be achieved by taking the target space of the theory to be 10D. If $c \neq 0$ the quantum theory is problematic to define, and we speak of a ‘conformal anomaly’ [Vonk (2005)].

The whole above story repeats itself for $\bar{T}(\bar{z})$ and its modes \bar{L}_m . At this point there is a crucial difference between open and closed strings. On an open string, left-moving and right-moving vibrations are related in such a way that they combine into standing waves. In our complex notation, ‘left-moving’ translates into ‘ z -dependent’ (i.e. holomorphic), and ‘right-moving’ into ‘ \bar{z} -dependent’ (i.e. anti-holomorphic). Thus, on an open string all holomorphic quantities are related to their anti-holomorphic counterparts. In particular, $T(z)$ and $\bar{T}(\bar{z})$, and their modes L_m and \bar{L}_m , turn out to be complex conjugates. There is therefore only one independent algebra of Virasoro generators L_m .

On a *closed string* on the other hand, which is the situation we have been studying so far, left- and right-moving waves are completely independent. This means that all holomorphic and anti-holomorphic quantities, and in particular $T(z)$ and $\bar{T}(\bar{z})$, are independent. One therefore has two sets of Virasoro generators, L_m and \bar{L}_m .

Let us now analyze the problem of central charge in the twisted theories. To twist the theory, we have used the $U(1)$ -symmetries. Any global $U(1)$ -symmetry of our theory has a conserved current J_α . The fact that it is conserved again means that $J_z \equiv J(z)$ is holomorphic and $J_{\bar{z}} \equiv \bar{J}(\bar{z})$

is anti-holomorphic. Once again, on an open string J and \bar{J} will be related, but in the closed string theory we are studying they will be independent functions. In particular, this means that we can view a global $U(1)$ -symmetry as really consisting of two independent, left- and right-moving, $U(1)$ -symmetries, with generators F_L and F_R .

Note that the sum of $U(1)$ -symmetries $F_V + F_A$ only acts on objects with a $+$ index. That is, it acts purely on left-moving quantities. Similarly, $F_V - F_A$ acts purely on right-moving quantities. From our discussion above, it is therefore natural to identify these two symmetries with the two components of a single global $U(1)$ symmetry:

$$F_V = \frac{1}{2}(F_L + F_R) \quad F_A = \frac{1}{2}(F_L - F_R).$$

A more detailed construction shows that this can indeed be done.

Let us expand the left-moving conserved $U(1)$ -current into Laurent modes,

$$J(z) = \sum J_m z^{-m-1}. \quad (6.322)$$

The commutation relations of these modes with one another and with the Virasoro modes can be calculated, either by writing down all of the modes in terms of the fields of the theory, or by using more abstract knowledge from the theory of superconformal symmetry algebras. In either case, one finds

$$\begin{aligned} [L_m, L_n] &= (m-n)L_{m+n} + \frac{c}{12}m(m^2-1)\delta_{m+n}[L_m, J_n] \\ &= -nJ_{m+n}[J_m, J_n] = \frac{c}{3}m\delta_{m+n}. \end{aligned}$$

Note that the same central charge c appears in the J - and in the L -commutators. This turns out to be crucial.

Following the standard Noether procedure, we can now construct a conserved charge by integrating the conserved current $J(z)$ over a space-like slice of the z -plane. In string theory, the physical time direction is the radial direction in the z -plane, so a space-like slice is just a curve around the origin. The integral is therefore calculated using the *Cauchy Theorem*,

$$F_L = \oint_{z=0} J(z)dz = 2\pi i J_0.$$

In the quantum theory, it will be this operator that generates the $U(1)_L$ -symmetry. Now recall that to twist the theory we want to introduce

new Lorentz rotation generators,

$$M_A = M - F_V = M - \frac{1}{2}(F_L + F_R)M_B = M - F_A = M - \frac{1}{2}(F_L - F_R).$$

A well-known result from string theory (see [Vonk (2005)]) is that the generator of Lorentz rotations is $M = 2\pi i(L_0 - \bar{L}_0)$. Therefore, we find that the twisting procedure in this new language amounts to

$$\begin{aligned} A : L_{0,A} &= L_0 - \frac{1}{2}J_0, & \bar{L}_{0,A} &= \bar{L}_0 + \frac{1}{2}\bar{J}_0, \\ B : L_{0,B} &= L_0 - \frac{1}{2}J_0, & \bar{L}_{0,B} &= \bar{L}_0 - \frac{1}{2}\bar{J}_0. \end{aligned}$$

Let us now focus on the left-moving sector; we see that for both twistings the new Lorentz rotation generator is the difference of L_0 and $\frac{1}{2}J_0$. The new Lorentz generator should also correspond to a conserved 2-tensor, and from (6.321) and (6.322) there is a very natural way to get such a current:

$$\tilde{T}(z) = T(z) + \frac{1}{2}\partial J(z), \quad (6.323)$$

which clearly satisfies $\bar{\partial}\tilde{T} = 0$ and

$$\tilde{L}_m = L_m - \frac{1}{2}(m+1)J_m, \quad (6.324)$$

so in particular we find that \tilde{L}_0 can serve as $L_{0,A}$ or $L_{0,B}$. We should apply the same procedure (with a minus sign in the A -model case) in the right-moving sector. Equations (6.323) and (6.324) tell us how to implement the twisting procedure not only on the conserved charges, but on the whole $N = 2$ superconformal algebra – or at least on the part consisting of the J - and L -modes, but a further investigation shows that this is the only part that changes. We have motivated, but not rigorously derived (6.323); for a complete justification the reader is referred to the original papers [Lerche *et. al.* (1989)] and [Cecotti and Vafa (1991)].

Now, we come to the crucial point. The algebra that the new modes \tilde{L}_m satisfy can be directly calculated from (6.323), and we find

$$[\tilde{L}_m, \tilde{L}_n] = (m - n)\tilde{L}_{m+n}.$$

That is, there is no central charge left. This means that we do not have any restriction on the dimension of the theory, and topological strings will actually be well-defined in target spaces of any dimension.

From this result, we see that we can integrate our partition function over conformally equivalent metrics without having to worry about the conformal anomaly represented by the nonzero central charge. After having integrated over this large part of the space of all metrics, it turns out that there is a nD integral left to do. In particular, it is known that one can always find a conformal transformation which in the neighborhood of a chosen point puts the metric in the form $h_{\alpha\beta} = \eta_{\alpha\beta}$, with η the usual flat metric with diagonal entries ± 1 . (Or, $+1$ in the Euclidean setting.) On the other hand, when one considers the global situation, it turns out that one cannot always enforce this gauge condition everywhere. For example, if the world-sheet is a torus, there is a left-over complex parameter τ that cannot be gauged away. The easiest way to visualize this parameter (see [Vonk (2005)]) is by drawing the resulting torus in the complex plane and rescaling it in such a way that one of its edges runs from 0 to 1; the other edge then runs from 0 to τ , see Figure 6.26. It seems intuitively clear that a conformal transformation – which should leave all angles fixed – will never deform τ , and even though intuition often fails when considering conformal mappings, in this case this can indeed be proven. Thus, τ is really a modular parameter which we need to integrate over. Another fairly intuitive result is that any locally flat torus can, after a rescaling, be drawn in this form, so τ indeed is the only modulus of the torus.

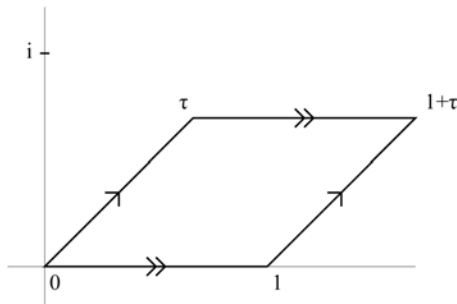


Fig. 6.26 The only modulus τ of a torus T^2 .

More generally, one can show that a Riemann surface of genus g has $m_g = 3(g - 1)$ complex modular parameters. As usual, this is the *virtual* dimension of the moduli space. If $g > 1$, one can show that this virtual dimension equals the actual dimension. For $g = 0$, the sphere, we have a negative virtual dimension $m_g = -3$, but the actual dimension is 0: there

is always a flat metric on a surface which is topologically a sphere (just consider the sphere as a plane with a point added at infinity), and after having chosen this metric there are no remaining parameters such as τ in the torus case. For $g = 1$, the virtual dimension is $m_g = 0$, but as we have seen the actual dimension is 1.

We can explain these discrepancies using the fact that, after we have used the conformal invariance to fix the metric to be flat, the sphere and the torus have leftover symmetries. In the case of the sphere, it is well known in string theory that one can use these extra symmetries to fix the positions of three labelled points. In the case of the torus, after fixing the metric to be flat we still have rigid translations of the torus left, which we can use to fix the position of a single labelled point. To see how this leads to a difference between the virtual and the actual dimensions, let us for example consider tori with n labelled points on them. Since the virtual dimension of the moduli space of tori without labelled points is 0, the virtual dimension of the moduli space of tori with n labelled points is n . One may expect that at some point (and in fact, this happens already when $n = 1$), one reaches a sufficiently generic situation where the virtual dimension really is the actual dimension. However, even in this case we can fix one of the positions using the remaining conformal (translational) symmetry, so the positions of the points only represent $n - 1$ moduli. Hence, there must be an n th modulus of a different kind, which is exactly the shape parameter τ that we have encountered above. In the limiting case where $n = 0$, this parameter survives, thus causing the difference between the virtual and the real dimension of the moduli space.

For the sphere, the reasoning is somewhat more formal: we analogously expect to have three ‘extra’ moduli when $n = 0$. In fact, three extra parameters are present, but they do not show up as moduli. They must be viewed as the three parameters which need to be added to the problem to find a 0D moduli space.

Since the cases $g = 0, 1$ are thus somewhat special, let us begin by studying the theory on a Riemann surface with $g > 1$. To arrive at the topological string correlation functions, after gauge fixing we have to integrate over the remaining moduli space of complex dimension $3(g - 1)$. To do this, we need to fix a measure on this moduli space. That is, given a set of $6(g - 1)$ tangent vectors to the moduli space, we want to produce a number which represents the size of the volume element spanned by these vectors, see Figure 6.27. We should do this in a way which is invariant under coordinate redefinitions of both the moduli space and the world-sheet.

Is there a ‘natural’ way to do this?

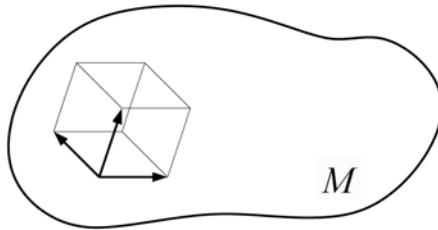


Fig. 6.27 A measure on the moduli space M assigns a number to every set of three tangent vectors. This number is interpreted as the volume of the element spanned by these vectors.

To answer this question, let us first ask how we can describe the tangent vectors to the moduli space (see [Vonk (2005)]). In two dimensions, conformal transformations are equivalent to holomorphic transformations: $z \mapsto f(z)$. It thus seems natural to assume that the moduli space we have left labels different complex structures on Σ , and indeed this can be shown to be the case. Therefore, a tangent vector to the moduli space is an infinitesimal change of complex structure, and these changes can be parameterized by holomorphic 1-forms with anti-holomorphic vector indices,

$$dz \mapsto dz + \epsilon \mu_{\bar{z}}^z(z) d\bar{z}.$$

The dimension counting above tells us that there are $3(g-1)$ independent $(\mu_i)_{\bar{z}}^z$, plus their $3(g-1)$ complex conjugates which change $d\bar{z}$. So the tangent space is spanned by these $\mu_i(z, \bar{z})$, $\bar{\mu}_i(z, \bar{z})$. How do we get a number out of a set of these objects? Since μ_i has a z and a \bar{z} index, it seems natural to integrate it over Σ . However, the z -index is an upper index, so we need to lower it first with some tensor with two z -indices. It turns out that a good choice is to use the Q -partner G_{zz} of the SEM-tensor component T_{zz} , and thus to define the integration over moduli space as

$$\int_{M_g} \prod_{i=1}^{3g=3} \left(dm^i d\bar{m}^i \int_{\Sigma} G_{zz}(\mu_i)_{\bar{z}}^z \int_{\Sigma} G_{\bar{z}\bar{z}}(\bar{\mu}_i)_{\bar{z}}^{\bar{z}} \right). \quad (6.325)$$

Note that by construction, this integral is also invariant under a change of basis of the moduli space. There are several reasons why using G_{zz} is a natural choice. First of all, this choice is analogous to what one does in bosonic string theory. There, one integrates over the moduli space using

exactly the same formula, but with G replaced by the conformal ghost b . This ghost is the BRST-partner of the SEM-tensor in exactly the same way as G is the Q -partner of T . Secondly, one can make the not unrelated observation that since $\{Q, G\} = T$, we can still use the standard arguments to show independence of the theory of the parameters in a Lagrangian of the form $L = \{Q, V\}$. The only difference is that now we also have to commute Q through G to make it act on the vacuum, but since $T_{\alpha\beta}$ itself is the derivative of the action with respect to the metric $h_{\alpha\beta}$, the terms we get in this way amount to integrating a total derivative over the moduli space. Therefore, apart from possible boundary terms these contributions vanish. Note that this reasoning also gives us an argument for using G_{zz} instead of T_{zz} (which is more or less the only other reasonable option) in (6.325): if we had chosen T_{zz} then all path integrals would have been over total derivatives on the moduli space, and apart from boundary contributions the whole theory would have become trivial.

If we consider the vector and axial charges of the full path integral measure, including the new path integral over the world-sheet metric h , we find a surprising result. Since the world-sheet metric does not transform under R -symmetry, naively one might expect that its measure does not either. However, this is clearly not correct since one should also take into account the explicit G -insertions in (6.325) that do transform under R -symmetry. From the $N = 2$ superconformal algebra (or, more down-to-earth, from expressing the operators in terms of the fields), it follows that the product of G and \bar{G} has vector charge zero and axial charge 2. Therefore, the total vector charge of the measure remains zero, and the axial charge gets an extra contribution of $6(g - 1)$, so we find a total axial R -charge of $6(g - 1) - 2m(g - 1)$. From this, we see that the case of complex target space dimension 3 is very special: here, the axial charge of the measure vanishes for any g , and hence the partition function is nonzero at every genus. If $m > 3$ and $g > 1$, the total axial charge of the measure is negative, and we have seen that we cannot cancel such a charge with local operators. Therefore, for these theories only the partition function at $g = 1$ and a specific set of correlation functions at genus zero give nonzero results. Moreover, for $m = 2$ and $m = 1$, the results can be shown to be trivial by other arguments. Therefore, a Calabi-Yau threefold is by far the most interesting target space for a topological string theory. It is a ‘happy coincidence’ (see [Vonk (2005)]) that this is exactly the dimension we are most interested in from the string theory perspective.

Finally, let us come back to the special cases of genus 0 and 1. At genus

zero, the Riemann surface has a single point as its moduli space, so there are no extra integrals or G -insertions to worry about. Therefore, we can copy the topological field theory result saying that we have to introduce local operators with total degree (m, m) in the theory. The only remnant of the fact that we are integrating over metrics is that we should also somehow fix the remaining three symmetries of the sphere. The most straightforward way to do this is to consider 3-point functions with insertions on three labelled points. As a gauge choice, we can then for example require these points to be at the points 0, 1 and ∞ in the compactified complex plane. For example, in the A -model on a Calabi–Yau threefold, the 3-point function of three operators corresponding to $(1, 1)$ -forms would thus give a nonzero result.

In the case of the torus, we have seen that there is one ‘unexpected’ modular parameter over which we have to integrate. This means we have to insert one G - and one \tilde{G} -operator in the measure, which spoils the absence of the axial anomaly we had for $g = 1$ in the topological field theory case. However, we also must fix the one remaining translational symmetry, which we can do by inserting a local operator at a labelled point. Thus, we can restore the axial R -charge to its zero value by choosing this to be an operator of degree $(1, 1)$.

Summarizing, we have found that in topological string theory on a target Calabi–Yau 3-fold, we have a non-vanishing 3-point function of total degree $(3, 3)$ at genus zero; a non-vanishing 1-point function of degree $(1, 1)$ at genus one, and a non-vanishing partition (‘zero-point’) function at all genera $g > 1$.

Nonlocal Operators

In one respect, what we have achieved is great progress: we can now for any genus define a nonzero partition function (or for low genus a correlation function) of the topological string theory. On the other hand, we would also like to define correlation functions of an arbitrary number of operators at these genera. As we have seen, the insertion of extra local operators in the correlation functions is not possible, since any such insertion will spoil our carefully constructed absence of R -symmetry anomalies. Therefore, we have to introduce nonlocal operators.

There is one class of nonlocal operators which immediately becomes mind. Before we saw, using the descent equations, that for every local operator we can define a corresponding 1-form and a 2-form operator. If

we check the axial and vector charges of these operators, we find that if we start with an operator of degree $(1, 1)$, the 2-form operator we end up with actually has vanishing axial and vector charges. This has two important consequences. First of all, we can add the integral of this operator to our action [Vonk (2005)],

$$S[t] = S_0 + t^a \int \mathcal{O}_a^{(2)},$$

without spoiling the axial and vector symmetry of the theory. Secondly, we can insert the integrated operator into correlation functions,

$$\langle \int \mathcal{O}_1^{(2)} \dots \int \mathcal{O}_n^{(2)} \rangle$$

and still get a nonzero result by the vanishing of the axial and vector charges. These two statements are related: one obtains such correlators by differentiating $S[t]$ with respect to the appropriate t 's, and then setting all $t^a = 0$.

A few remarks are in place here. First of all, recall that the integration over the insertion points of the operators can be viewed as part of the integration over the moduli space of Riemann surfaces, where now we label a certain number of points on the Riemann surface. From this point of view, the $g = 0, 1$ cases fit naturally into the same framework. We could unite the descendant fields into a world-sheet *super-field*,

$$\Phi_a = \mathcal{O}_a^{(0)} + \mathcal{O}_{a\alpha}^{(1)} \theta^\alpha + \mathcal{O}_{a\alpha\beta}^{(2)} \theta^\alpha \theta^\beta$$

where we formally replaced each dz and $d\bar{z}$ by corresponding fermionic coordinates θ^z and $\theta^{\bar{z}}$. Now, one can write the above correlators as integrals over n copies of this *super-space*,

$$\int \prod_{s=1}^n d^2 z_s d^2 \theta_s \langle \Phi_{a_1}(z_1, \theta_1) \dots \Phi_{a_n}(z_n, \theta_n) \rangle$$

The integration prescription at genus 0 and 1 tells us to fix 3 and 1 points respectively, so we need to remove this number of super-space integrals. Then, integrating over the other super-space coordinates, the genus 0 correlators indeed become

$$\langle \mathcal{O}_{a_1}^{(0)} \mathcal{O}_{a_2}^{(0)} \mathcal{O}_{a_3}^{(0)} \int \mathcal{O}_{a_4}^{(2)} \dots \int \mathcal{O}_{a_n}^{(2)} \rangle$$

From this prescription we note that these expressions are symmetric in the exchange of all a_i and a_j . In particular, this means that the genus zero

3-point functions at arbitrary t ,

$$c_{abc}[t] = \langle \mathcal{O}_a^{(0)} \mathcal{O}_b^{(0)} \mathcal{O}_c^{(0)} \rangle [t]$$

have symmetric derivatives:

$$\frac{\partial c_{abc}}{\partial t^d} = \frac{\partial c_{abd}}{\partial t^c},$$

and similarly with permuted indices. These equations can be viewed as integrability conditions, and using the Poincaré lemma we see that they imply that

$$c_{ijk}[t] = \frac{\partial Z_0[t]}{\partial t^i \partial t^j \partial t^k}.$$

for some function $Z_0[t]$. Following the general philosophy that n -point functions are n th derivatives of the t -dependent partition function, we see that $Z_0[t]$ can be naturally thought of as the partition function at genus zero. Similarly, the partition function at genus 1 can be defined by integrating up the one-point functions once.

The quantities we have calculated above should be semi-topological invariants, meaning that they only depend on ‘half’ of the moduli (either the Kähler ones or the complex structure ones) of the target space. For example, in the A -model we find the Gromov–Witten invariants. In the B -model, it turns out that $F_0[t] = \ln Z_0[t]$ is actually a quantity we already knew: it is the prepotential of the Calabi–Yau manifold. A discussion of why this is the case can be found in the paper [Bershadsky *et. al.* (1994)]. The higher genus partition functions can be thought of as ‘quantum corrections’ to the prepotential.

Finally, there is a type of operator we have not discussed at all so far. Recall that in the topological string theory, the metric itself is now a dynamical field. We could not include the metric in our physical operators, since this would spoil the topological invariance. However, the metric is part of a Q -multiplet, and the highest field in this multiplet is a scalar field which is usually labelled φ . (It should not be confused with the fields ϕ^i .) We can get more correlation functions by inserting operators φ^k and the operators related to them by the descent equations into the correlation functions. These operators are called ‘gravitational descendants’. Even the case where the power is $k = 0$ is nontrivial; it does not insert any operator, but it does label a certain point, and hence changes the moduli space one integrates over. This operator is called the ‘puncture operator’.

All of this seems to lead to an enormous amount of semi-topological target space invariants that can be calculated, but there are many recursion relations between the several correlators. This is similar to how we showed before that all correlators for the cohomological field theories follow from the 2- and 3-point functions on the sphere. Here, it turns out that the set of all correlators has a structure which is related to the theory of integrable hierarchies. Unfortunately, a discussion of this is outside the scope of both these lectures and the author's current knowledge.

The Holomorphic Anomaly

We have now defined the partition function and correlation functions of topological string theory, but even though the expressions we obtained are much simpler than the path integrals for ordinary quantum field or string theories, it would still be very hard to explicitly calculate them. Fortunately, it turns out that the t -dependent partition and correlation functions are actually ‘nearly holomorphic’ in t , and this is a great aid in exactly calculating these quantities.

Let us make this ‘near holomorphy’ more precise. As we have seen, calculating correlation functions of primary operators in topological string theories amounts to taking t -derivatives of the corresponding perturbed partition function $Z[t]$ and consequently setting $t = 0$. Recall that $Z[t]$ is defined through adding terms to the action of the form

$$t^a \int_{\Sigma} \mathcal{O}_a^{(2)}, \quad (6.326)$$

Let us for definiteness consider the B -twisted model. We want to show that the above term is Q_B -exact. For simplicity, we assume that $\mathcal{O}_a^{(2)}$ is a bosonic operator, but what we are about to say can be generalized to the fermionic case. From the descent equations we studied in the subsection 6.7.4 above, we know that

$$(\mathcal{O}_a^{(2)})_{+-} = -\{G_+, [\mathcal{O}_a^{(0)}]\}, \quad (6.327)$$

where G_+ is the charge corresponding to the current G_{zz} , and G_- the one corresponding to $G_{\bar{z}\bar{z}}$. We can in fact express G_{\pm} in terms of the $N = (2, 2)$ supercharges \mathcal{Q} . So, according to [Vonk (2005)], we have

$$\begin{aligned} H = 2\pi i(L_0 + \bar{L}_0) &= \frac{1}{2}\{\mathcal{Q}_+, \bar{\mathcal{Q}}_+\} - \frac{1}{2}\{\mathcal{Q}_-, \bar{\mathcal{Q}}_-\}P \\ &= 2\pi i(L_0 - \bar{L}_0) = \frac{1}{2}\{\mathcal{Q}_+, \bar{\mathcal{Q}}_+\} + \frac{1}{2}\{\mathcal{Q}_-, \bar{\mathcal{Q}}_-\}. \end{aligned}$$

Thus, we find that the left- and right-moving SEM charges satisfy

$$T_+ = 2\pi i L_0 = \frac{1}{2}\{\mathcal{Q}_+, \bar{\mathcal{Q}}_+\}T_- = 2\pi i \bar{L}_0 = -\frac{1}{2}\{\mathcal{Q}_-, \bar{\mathcal{Q}}_-\}.$$

To find G in the B -model, we should write these charges as commutators with respect to $Q_B = \bar{\mathcal{Q}}_+ + \bar{\mathcal{Q}}_-$, which gives

$$T_+ = \frac{1}{2}\{Q_B, \mathcal{Q}_+\}T_- = -\frac{1}{2}\{Q_B, \mathcal{Q}_-\},$$

so we arrive at the conclusion that for the B -model,

$$G_+ = \frac{1}{2}\mathcal{Q}_+G_- = -\frac{1}{2}\mathcal{Q}_-.$$

Now, we can rewrite (6.327) as

$$\begin{aligned} (\mathcal{O}_a^{(2)})_{+-} &= -\{G_+, [G_-, \mathcal{O}_a^{(0)}]\} = \frac{1}{4}\{\mathcal{Q}_+, [\mathcal{Q}_-, \mathcal{O}_a^{(0)}]\} \\ &= \frac{1}{8}\{\bar{\mathcal{Q}}_B, [(\mathcal{Q}_- - \mathcal{Q}_+), \mathcal{O}_a^{(0)}]\}, \end{aligned} \quad (6.328)$$

which proves our claim that $\mathcal{O}_a^{(2)}$ is Q_B -exact.

An $N = (2, 2)$ sigma model with a real action does, apart from the term (6.326), also contain a term

$$t^{\bar{a}} \int_{\Sigma} \bar{\mathcal{O}}_a^{(2)}, \quad (6.329)$$

where $t^{\bar{a}}$ is the complex conjugate of t^a . It is not immediately clear that $\bar{\mathcal{O}}_a^{(2)}$ is a physical operator: we have seen that physical operators in the B -model correspond to forms that are $\bar{\partial}$ -closed, but the complex conjugate of such a form is ∂ -closed. However, by taking the complex conjugate of (6.328), we see that

$$(\bar{\mathcal{O}}_a^{(2)})_{+-} = \frac{1}{8}\{Q_B, [(\bar{\mathcal{Q}}_- - \bar{\mathcal{Q}}_+), \bar{\mathcal{O}}_a^{(0)}]\},$$

so not only is the operator Q_B -closed, it is even Q_B -exact. This means that we can add terms of the form (6.329) to the action, and taking $t^{\bar{a}}$ -derivatives inserts Q_B -exact terms in the correlation functions. Naively, we would expect this to give a zero result, so all the physical quantities seem to be t -independent, and thus holomorphic in t . We will see in a moment that this naive expectation turns out to be almost right, but not quite.

However, before doing so, let us comment briefly on the generalization of the above argument in the case of the A -model. It seems that a straightforward generalization of the argument fails, since Q_A is its own complex conjugate, and the complex conjugate of the de Rham operator is also the same operator. However, note that the $N = (2, 2)$ -theory has a different kind of ‘conjugation symmetry’: we can exchange the two supersymmetries, or in other words, exchange θ^+ with $\bar{\theta}^+$ and θ^- with $\bar{\theta}^-$. This exchanges Q_A with an operator which we might denote as $Q_{\bar{A}} \equiv Q_+ + \bar{Q}_-$. Using the above argument, we then find that the physical operators $\mathcal{O}_a^{(2)}$ are $Q_{\bar{A}}$ -exact, and that their conjugates in the new sense are Q_A -exact. We can now add these conjugates to the action with parameters $t^{\bar{a}}$, and we again naively find independence of these parameters. In this case it is less natural to choose t^a and $t^{\bar{a}}$ to be complex conjugates, but we are free to choose this particular ‘background point’ and study how the theory behaves if we then vary t^a and $t^{\bar{a}}$ independently.

Now, let us see how the naive argument showing independence of the theory of $t^{\bar{a}}$ fails. In fact, the argument above would certainly hold for topological *field* theories. However, in topological string theories (see [Vonk (2005)]), we have to worry about the insertions in the path integral of

$$G \cdot \mu_i \equiv \int d^2 z G_{zz} (\mu_i)_{\bar{z}},$$

and their complex conjugates, when commuting the Q_B towards the vacuum and making sure it gives a zero answer. Indeed, the Q_B -commutator of the above factor is not zero, but it gives

$$\{Q_B, G \cdot \mu_i\} = T \cdot \mu_i.$$

Now recall that $T_{\alpha\beta} = \partial_{h^{\alpha\beta}} S$. We did not give a very precise definition of μ_i above, but we know that it parameterizes the change in the metric under an infinitesimal change of the coordinates m_i on the moduli space. One can make this intuition precise, and then finds the following ‘chain rule’: $T \cdot \mu_i = \partial_{m^i} S$. Inserting this into the partition function, we find that

$$\frac{\partial F_g}{\partial t^{\bar{a}}} = \int_{M_g} \prod_{i=1}^{3g-3} dm^i d\bar{m}^i \sum_{j,k} \frac{\partial^2}{\partial m^j \partial \bar{m}^k} \left\langle \left(\prod_{l \neq j} \int \mu_l \cdot G \right) \left(\prod_{l \neq k} \int \bar{\mu}_l \cdot \bar{G} \right) \int \bar{\mathcal{O}}_a^{(2)} \right\rangle,$$

where $F_g = \ln Z_g$ is the free energy at genus g , and the reason F_g appears in the above equation instead of Z_g is, as usual in quantum field theory, that the expectation values in the r.h.s. are normalized such that $\langle 1 \rangle = 1$, and so the l.h.s. should be normalized accordingly and equal $Z_g^{-1} \partial_{\bar{a}} Z_g = \partial_{\bar{a}} F_g$ [Vonk (2005)].

Thus, as we have claimed before, we are integrating a total derivative over the moduli space of genus g surfaces. If the moduli space did not have a boundary, this would indeed give zero, but in fact the moduli space does have a boundary. It consists of the moduli which make the genus g surface degenerate. This can happen in two ways: an internal cycle of the genus g surface can be pinched, leaving a single surface of genus $g - 1$, as in Figure 6.28 (a), or the surface can split up into two surfaces of genus g_1 and $g_2 = g - g_1$, as depicted in Figure 6.28 (b). By carefully considering the boundary contributions to the integral for these two types of boundaries, it was shown in [Bershadsky *et. al.* (1994)] that

$$\frac{\partial F_g}{\partial t^{\bar{a}}} = \frac{1}{2} c_{\bar{a}\bar{b}\bar{c}} e^{2K} G^{\bar{b}d} G^{\bar{c}e} \left(D_d D_e F_{g-1} + \sum_{r=1}^{g-1} D_d F_r D_e F_{g-r} \right),$$

where G is the so-called *Zamolodchikov metric* on the space parameterized by the coupling constants $t^a, t^{\bar{a}}$; K is its Kähler potential, and the D_a are covariant derivatives on this space. The coefficients $c_{\bar{a}\bar{b}\bar{c}}$ are the 3-point functions on the sphere of the operators $\bar{\mathcal{O}}_a^{(0)}$. We will not derive the above formula in detail, but the reader should notice that the contributions from the two types of boundary are quite clear.

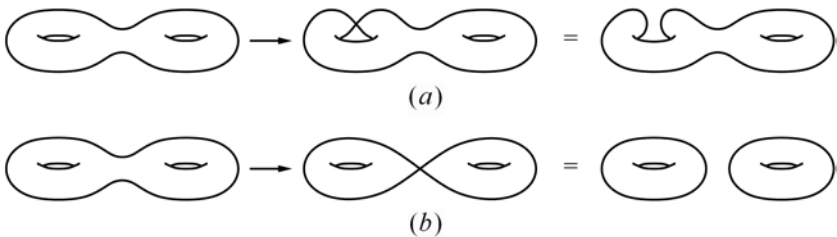


Fig. 6.28 At the boundary of the moduli space of genus g surfaces, the surfaces degenerate because certain cycles are pinched. This either lowers the genus of the surface (a) or breaks the surface into two lower genus ones (b) (see text for explanation).

Using this formula, one can inductively determine the $t^{\bar{a}}$ dependence on the partition functions if the holomorphic t^a -dependence is known.

Holomorphic functions on complex spaces (or more generally holomorphic sections of complex vector bundles) are quite rare: usually, there is only a nD space of such functions. The same turns out to hold for our topological string partition functions: even though they are not quite holomorphic, their anti-holomorphic behavior is determined by the holomorphic dependence on the coordinates, and as a result there is a finite number of coefficients which determines them.

Thus, just from the above structure and without doing any path integrals, one can already determine the topological string partition functions up to a finite number of constants. This leads to a feasible program for completely determining the topological string partition function for a given target space and at given genus. From the holomorphic anomaly equation, one first has to find the general form of the partition function. Then, all one has left to do is to fix the unknown constants. Here, the fact that in the A -model the partition function counts a number of points becomes our help: by requiring that the A -model partition functions are integral, one can often fix the unknown constants and completely determine the t -dependent partition function. In practice, the procedure is still quite elaborate, so we will not describe any examples here, but several have been worked out in detail in the literature. Once again, the pioneering work for this can be found in the paper [Bershadsky *et. al.* (1994)].

6.7.6 Geometrical Transitions

Conifolds

Recall that a *conifold* is a generalization of the notion of a manifold. Unlike manifolds, a conifold can (or, should) contain conical singularities i.e., points whose neighborhood looks like a cone with a certain base. The base is usually a 5D manifold.

In string theory, a *conifold transition* represents such an evolution of the Calabi–Yau manifold in which its fabric rips and repairs itself, yet with mild and acceptable physical consequences in the context of string theory. However, the tears involved are more severe than those in an ‘weaker’ *flop transition* (see [Greene (2000)]). The geometrically singular conifolds were shown to lead to completely smooth physics of strings. The divergences are ‘smeared out’ by D3-branes wrapped on the shrinking 3-sphere S^3 , as originally pointed out by A. Strominger, who, together with D. Morrison

and B. Greene have also found that the topology near the conifold singularity can undergo a topological phase-transition (see subsection 6.4.5). It is believed that nearly all Calabi–Yau manifolds can be connected via these ‘critical transitions’.

More precisely, the conifold is the simplest example of a non-compact Calabi–Yau 3-fold: it is the set of solutions to the equation

$$x_1x_2 - x_3x_4 = 0$$

in \mathbb{C}^4 . The resulting manifold is a cone, meaning in this case that any real multiple of a solution to this equation is again a solution. The point $(0, 0, 0, 0)$ is the ‘tip’ of this cone, and it is a singular point of the solution space. Note that by writing

$$x_1 = z_1 + iz_2, \quad x_2 = z_1 - iz_2, \quad x_3 = z_3 + iz_4, \quad x_4 = -z_3 + iz_4,$$

where the z_i are still *complex* numbers, one can also write the equation as

$$z_1^2 + z_2^2 + z_3^2 + z_4^2 = 0.$$

Writing each z_i as $a_i + ib_i$, with a_i and b_i real, we get the two equations

$$|a|^2 - |b|^2 = 0, \quad a \cdot b = 0. \quad (6.330)$$

Here $a \cdot b = \sum_i a_i b_i$ and $|a|^2 = a \cdot a$. Since the geometry is a cone, let us focus on a ‘slice’ of this cone given by

$$|a|^2 + |b|^2 = 2r^2,$$

for some $r \in \mathbb{R}$. On this slice, the first equation in (6.330) becomes

$$|a|^2 = r^2, \quad (6.331)$$

which is the equation defining a 3-sphere S^3 of radius r . The same holds for b , so both a and b lie on 3-spheres. However, we also have to take the second equation in (6.330) into account. Let us suppose that we fix an a satisfying (6.331). Then b has to lie on a 3-sphere, but also on the plane through the origin defined by $a \cdot b = 0$. That is, b lies on a 2-sphere. This holds for every a , so the slice we are considering is a fibration of 2-spheres over the 3-sphere. With a little more work, one can show that this fibration is trivial, so the conifold is a cone over $S^2 \times S^3$.

Since the conifold is a singular geometry, we would like to find geometries which approximate it, but which are non-singular. There are two

interesting ways in which this can be done. The simplest way is to replace the defining equation by

$$x_1x_2 - x_3x_4 = \mu^2. \quad (6.332)$$

From the two equations constraining a and b , we now see that $|a|^2 \geq \mu^2$. In other words, the parameter r should be at least μ . At $r = \mu$, the a -sphere still has finite radius μ , but the b -sphere shrinks to zero size. This geometry is called the *deformed conifold*. Even though this is not clear from the picture, from the equation (6.332) one can straightforwardly show that it is nonsingular. One can also show that it is topologically equivalent to the cotangent bundle on the 3-sphere, T^*S^3 . Here, the S^3 on which the cotangent bundle is defined is exactly the S^3 at the ‘tip’ of the deformed conifold.

The second way to change the conifold geometry arises from studying the two equations

$$x_1A + x_3B = 0, \quad x_4A + x_2B = 0. \quad (6.333)$$

Here, we require A and B to be homogeneous complex coordinates on a \mathbb{CP}^1 , i.e.,

$$(A, B) \neq (0, 0), \quad (A, B) \sim (\lambda A, \lambda B)$$

where λ is any nonzero complex number. If one of the x_i is nonzero, say x_1 , one can solve for A or B , e.g., $A = -\frac{x_3B}{x_1}$, and insert this in the other equation to get

$$x_1x_2 - x_3x_4 = 0$$

which is the conifold equation. However, if all x_i are zero, any A and B solve the system of equations (6.333). In other words, we have constructed a geometry which away from the former singularity is completely the same as the conifold, but the singularity itself is replaced by a \mathbb{CP}^1 , which topologically is the same as an S^2 . From the defining equations one can again show that the resulting geometry is nonsingular, so we have now replaced our conifold geometry by the so-called *resolved conifold*.

Topological D-branes

Since topological string theories are in many ways similar to an ordinary (bosonic) string theories, one natural question which arises is: are there also open topological strings which can end on D-branes? To answer the

above question rigorously, we would have to study boundary conditions on world-sheets with boundaries which preserve the Q -symmetry.

In the A -model, one can only construct 3D-branes wrapping so-called ‘Lagrangian’ submanifolds of M . Here, ‘Lagrangian’ means that the Kähler form ω vanishes on this submanifold. In the B -model, one can construct D-branes of any even dimension, as long as these branes wrap *holomorphic* submanifolds of M .

Just like in ordinary string theory, when we consider open topological strings ending on a D-brane, there should be a field theory on the brane world-volume describing the low-energy physics of the open strings. Moreover, since we are studying topological theories, one may expect such a theory to inherit the property that it only depends on a restricted amount of data of the manifolds involved. A key example is the case of the A -model on the deformed conifold, $M = T^*S^3$, where we wrap ND -branes on the S^3 in the base. (One can show that this is indeed a Lagrangian submanifold.) In ordinary string theory, the world-volume theory on ND -branes has a $U(N)$ gauge symmetry, so putting the ingredients together we can make the guess that the world-volume theory is a 3D topological field theory with $U(N)$ gauge symmetry. There is really only one candidate for such a theory: the *Chern–Simons gauge theory*. Recall that it consists of a single $U(N)$ gauge field, and has the action

$$S = \frac{k}{4\pi} \int_{S^3} \text{Tr} \left(A \wedge dA + \frac{2}{3} A \wedge A \wedge A \right). \quad (6.334)$$

Before the invention of D-branes, E. Witten showed that this is indeed the theory one gets. In fact, he showed even more: this theory actually describes the *full* topological string-field theory on the D-branes, even without going to a low-energy limit [Witten (1995a)].

Let us briefly outline the argument that gives this result. In his paper, Witten derived the open string-field theory action for the open A -model topological string; it reads

$$S = \int \text{Tr} \left(\mathcal{A} * Q_A \mathcal{A} + \frac{2}{3} \mathcal{A} * \mathcal{A} * \mathcal{A} \right).$$

The form of this action is very similar to Chern–Simons theory, but its interpretation is completely different: \mathcal{A} is a *string-field* (a wave function on the space of all maps from an open string to the space-time manifold), Q_A is the topological symmetry generator, which has a natural action on the string-field, and $*$ is a certain noncommutative product. Witten shows that

the topological properties of the theory imply that only the constant maps contribute, so \mathcal{A} becomes a field on M – and since open strings can only end on D-branes, it actually becomes a field on S^3 . Moreover, recall that Q_A can be interpreted as a de Rham differential. Using these observations and the precise definition of the star product one can indeed show that the string-field theory action reduces to Chern–Simons theory on S^3 .

6.7.7 Topological Strings and Black Hole Attractors

Topological string theory is naturally related to *black hole dynamics* (see subsection 5.12.3 above). Namely, critical string theory compactified on Calabi–Yau manifolds has played a central role in both the mathematical and physical development of modern string theory. The physical relevance of the data provided by the topological string $\hat{c} = 6$ (of A and B types) has been that it computes F –type terms in the corresponding four dimensional theory [Bershadsky *et. al.* (1994); Antoniadis *et. al.* (1994)]. These higher-derivative F –type terms for Type II superstring on a Calabi–Yau manifold are of the general form

$$\int d^4x d^4\theta (W_{ab}W^{ab})^g F_g(X^\Lambda), \quad (6.335)$$

where W_{ab} is the *graviphoton super-field* of the $N = 2$ *super-gravity* and X^Λ are the *vector multiplet fields*. The lowest component of W is F the graviphoton field strength and the highest one is the Riemann tensor. The lowest components of X^Λ are the complex scalars parameterizing Calabi–Yau moduli and their highest components are the associated $U(1)$ vector-fields. These terms contribute to multiple graviphoton–graviton scattering. (6.335) includes (after θ integrations) an $R^2 F^{2g-2}$ term. The topological string partition function Z_{top} represents the canonical ensemble for multi-particle spinning five dimensional black holes [Breckenridge *et. al.* (1997); Katz *et. al.* (1999)].

Recently, [Ooguri *et. al.* (2004)] proposed a simple and direct relationship between the second-quantized topological string partition function Z_{top} and black hole partition function Z_{BH} in four dimensions of the form

$$Z_{BH}(p^\Lambda, \phi^\Lambda) = |Z_{\text{top}}(X^\Lambda)|^2, \quad \text{where} \quad X^\Lambda = p^\Lambda + \frac{i}{\pi} \phi^\Lambda$$

in a certain *Kähler gauge*. The l.h.s. here is evaluated as a function of integer magnetic charges p^Λ and continuous electric potentials ϕ^Λ , which are conjugate to integer electric charges q_Λ . The r.h.s. is the holomorphic

square of the partition function for a gas of topological strings on a Calabi–Yau whose moduli are those associated to the charges/potentials $(p^\Lambda, \phi^\Lambda)$ via the attractor equations [Ooguri *et. al.* (2004)]. Both sides of (6.336) are defined in a perturbation expansion in $1/Q$, where Q is the graviphoton charge carried by the black hole.²⁹ The non-perturbative completion of either side of (6.336) might in principle be defined as the partition function of the holographic CFT dual to the black hole, as in [Strominger and Vafa (1996)]. Then we have the triple equality,

$$Z_{CFT} = Z_{BH} = |Z_{\text{top}}|^2.$$

The existence of fundamental connection between 4D black holes and the topological string might have been anticipated from the following observation. Calabi–Yau spaces have two types of moduli: Kähler and complex structure. The world-sheet twisting which produces the A (B) model topological string from the critical superstring eliminates all dependence on the complex structure (Kähler) moduli at the perturbative level. Hence the perturbative topological string depends on only half the moduli. Black hole entropy on the other hand, insofar as it is an intrinsic property of the black hole, cannot depend on any externally specified moduli. What happens at leading order is that the moduli in vector multiplets are driven to attractor values at the horizon which depend only on the black hole charges and not on their asymptotically specified values. Hypermultiplet vevs on the other hand are not fixed by an attractor mechanism but simply drop out of the entropy formula. It is natural to assume this is valid to all orders in a $1/Q$ expansion. Hence the perturbative topological string and the large black hole partition functions depend on only half the Calabi–Yau moduli. It would be surprising if string theory produced two functions on the same space that were not simply related. Indeed [Ooguri *et. al.* (2004)] argued that they were simply related as in (6.336).

Supergravity Area–Entropy Formula

Recall that a well-known hypothesis by J. Bekenstein and S. Hawking states that *the entropy of a black hole is proportional to the area of its horizon* (see [Hawking and Israel (1979)]). This *area* is a function of the black hole mass, or in the extremal case, of its charges. Here we review the leading *semiclassical area–entropy formula* for a general $N = 2$, $d = 4$ extremal black hole characterized by magnetic and electric charges (p^Λ, q_Λ) , recently

²⁹The string coupling g_s is in a hypermultiplet and decouples from the computation.

reviewed in [Ooguri *et. al.* (2004)]. The asymptotic values of the moduli in vector multiplets, parameterized by complex projective coordinates X^Λ , ($\Lambda = 0, 1, \dots, n_V$) in the black hole solution, are arbitrary. These moduli couple to the electromagnetic fields and accordingly vary as a function of the radius. At the horizon they approach an attractor point whose location in the moduli space depends only on the charges. The locations of these attractor points can be found by looking for supersymmetric solutions with constant moduli. They are determined by the attractor equations,

$$p^\Lambda = \text{Re}[CX^\Lambda], \quad q_\Lambda = \text{Re}[CF_{0\Lambda}], \quad (6.336)$$

where $F_{0\Lambda} = \partial F_0 / \partial X^\Lambda$ are the *holomorphic periods*, and the subscript 0 distinguishes these from the string loop corrected periods to appear in the next subsection. Both (p^Λ, q_Λ) and $(X^\Lambda, F_{0\Lambda})$ transform as vectors under the $Sp(2n+2; Z)$ duality group.

The $(2n_v + 2)$ real equations (6.336) determine the $(n_v + 2)$ complex quantities (C, X^Λ) up to Kähler transformations, which act as

$$K \rightarrow K - f(X) - \bar{f}(\bar{X}), \quad X^\Lambda \rightarrow e^f X^\Lambda, \quad F_0 \rightarrow e^{2f} F_0, \quad C \rightarrow e^{-f} C,$$

where the *Kähler potential* K is given by

$$e^{-K} = i(\bar{X}^\Lambda F_{0\Lambda} - X^\Lambda \bar{F}_{0\Lambda}).$$

We could at this point set $C = 1$ and fix the Kähler gauge but later we shall find other gauges useful. It is easy to see that (as required) the charges (p^Λ, q_Λ) determined by the attractor equations (6.336) are invariant under Kähler transformations. Given the horizon attractor values of the moduli determined by (6.336) the *Bekenstein–Hawking entropy* S_{BH} may be written as

$$S_{BH} = \frac{1}{4} \text{Area} = \pi |Q|^2,$$

where $Q = Q_m + iQ_e$ is a complex combination of the magnetic and electric graviphoton charges and

$$|Q|^2 = \frac{i}{2} (q_\Lambda \bar{C} \bar{X}^\Lambda - p^\Lambda \bar{C} \bar{F}_{0\Lambda}) = \frac{C\bar{C}}{4} e^{-K}.$$

The normalization of Q here is chosen so that $|Q|$ equals the radius of the two sphere at the horizon.

It is useful to rephrase the above results in the context of type IIB superstrings in terms of geometry of Calabi–Yau. In this case the attractor equations fix the complex geometry of the Calabi–Yau. The electric/magnetic charges correlate with three cycles of Calabi–Yau. Choosing a symplectic basis for the three cycles gives a choice of the splitting to electric and magnetic charges. Let A_Λ denote a basis for the electric three cycles, B^Σ the dual basis for the magnetic charges and Ω the holomorphic 3-form at the attractor point. Ω is fixed up to an overall multiplication by a complex number $\Omega \rightarrow \lambda\Omega$. There is a unique choice of λ such that the resulting Ω has the property that

$$p^\Lambda = \int_{A_\Lambda} \text{Re } \Omega = \text{Re}[CX^\Lambda], \quad q_\Lambda = \int_{B^\Lambda} \text{Re } \Omega = \text{Re}[CF_{0\Lambda}],$$

$$\text{where } \text{Re } \Omega = \frac{1}{2}(\Omega + \bar{\Omega}).$$

In terms of this choice, the black hole entropy can be written as

$$S_{BH} = \frac{\pi}{4} \int_{CY} \Omega \wedge \bar{\Omega}.$$

Higher–Order Corrections

F -term corrections to the action are encoded in a string loop corrected holomorphic prepotential

$$F(X^\Lambda, W^2) = \sum_{h=0}^{\infty} F_h(X^\Lambda) W^{2h}, \quad (6.337)$$

where F_h can be computed by topological string amplitudes (as we review in the next section) and W^2 involves the square of the anti-self dual graviphoton field strength. This obeys the homogeneity equation

$$X^\Lambda \partial_\Lambda F(X^\Lambda, W^2) + W \partial_W F(X^\Lambda, W^2) = 2F(X^\Lambda, W^2). \quad (6.338)$$

Near the black hole horizon, the attractor value of W^2 obeys $C^2 W^2 = 256$, and therefore the exact attractor equations read

$$p^\Lambda = \text{Re}[CX^\Lambda], \quad q_\Lambda = \text{Re} \left[CF_\Lambda \left(X^\Lambda, \frac{256}{C^2} \right) \right]. \quad (6.339)$$

This is essentially the only possibility consistent with *symplectic invariance*. It has been then argued that the entropy as a function of the charges is

$$S_{BH} = \frac{\pi i}{2} (q_\Lambda \bar{C} \bar{X}^\Lambda - p^\Lambda \bar{C} \bar{F}_\Lambda) + \frac{\pi}{2} \text{Im}[C^3 \partial_C F], \quad (6.340)$$

where F_Λ , X^Λ and C are expressed in terms of the charges using (6.339).

Topological Strings

Partition Functions for Black Hole and Topological Strings. The notion of *topological string* was introduced in [Witten (1990)]. Subsequently a connection between them and superstring was discovered: It was shown in [Bershadsky *et. al.* (1994); Antoniadis *et. al.* (1994)], that the superstring loop corrected F -terms (6.337) can be computed as topological string amplitudes. The purpose of this subsection is to translate the super-gravity notation of the previous section to the topological string notation.

The second quantized partition function for the topological string may be written

$$\begin{aligned} Z_{\text{top}}(t^A, g_{\text{top}}) &= \exp [F_{\text{top}}(t^A, g_{\text{top}})], \quad \text{where} \\ F_{\text{top}}(t^A, g_{\text{top}}) &= \sum_h g_{\text{top}}^{2h-2} F_{\text{top},h}(t^A), \end{aligned}$$

and $F_{\text{top},h}$ is the h -loop *topological string amplitude*. The Kähler moduli are expressed in the flat coordinates

$$t^A = \frac{X^A}{X^0} = \theta^A + i r^A,$$

where r^A are the Kähler classes of the Calabi-Yau M and θ^A are periodic $\theta^A \sim \theta^A + 1$.

We would like to determine relations between super-gravity quantities and topological string quantities. Using the homogeneity property (6.338) and the expansion (6.337), the holomorphic prepotential in super-gravity can be expressed as

$$\begin{aligned} F(CX^\Lambda, 256) &= (CX^0)^2 F\left(\frac{X^\Lambda}{X^0}, \frac{256}{(CX^0)^2}\right) \\ &= \sum_{h=0}^{\infty} (CX^0)^{2-2h} f_h(t^A), \end{aligned} \quad (6.341)$$

where $f_h(t^A)$ is related to $F_h(X^\Lambda)$ in (6.337) as

$$f_h(t^A) = 16^{2h} F_h\left(\frac{X^\Lambda}{X^0}\right).$$

This suggests an identification of the form $f_h(t^A) \sim F_{\text{top},h}(t^A)$ and $g_{\text{top}} \sim (CX^0)^{-1}$. For later purposes, we need precise relations between super-gravity and topological string quantities, including numerical coefficients. These can be determined by studying the limit of a large Calabi–Yau space.

In the *super-gravity notation*, the genus 0 and 1 terms in the large volume are given by

$$\begin{aligned} F(CX^\Lambda, 256) &= C^2 D_{ABC} \frac{X^A X^B X^C}{X^0} - \frac{1}{6} c_{2A} \frac{X^A}{X^0} + \cdots \\ &= (CX^0)^2 D_{ABC} t^A t^B t^C - \frac{1}{6} c_{2A} t^A + \cdots, \\ \text{where } c_{2A} &= \int_M c_2 \wedge \alpha_A, \end{aligned}$$

with c_2 being the *second Chern class* of M , and $C_{ABC} = -6D_{ABC}$ are the 4-cycle intersection numbers. These terms are normalized so that the mixed entropy S_{BH} is given by (6.340). On the other hand, the topological string amplitude in this limit is given by

$$F_{\text{top}} = -\frac{(2\pi)^3 \mathbf{i}}{g_{\text{top}}^2} D_{ABC} t^A t^B t^C - \frac{\pi \mathbf{i}}{12} c_{2A} t^A + \cdots \quad (6.342)$$

The normalization here is fixed by the holomorphic anomaly equations in [Bershadsky *et. al.* (1994)], which are nonlinear equations for $F_{\text{top},h}$.

Comparing the one-loop terms in (6.341) and (6.342), which are independent of g_{top} , we find

$$F(CX^\Lambda, 256) = -\frac{2\mathbf{i}}{\pi} F_{\text{top}}(t^A, g_{\text{top}}).$$

Given this, we can compare the genus 0 terms to find

$$g_{\text{top}} = \pm \frac{4\pi \mathbf{i}}{CX^0}.$$

This implies

$$\begin{aligned} \ln Z_{BH} &= -\pi \operatorname{Im} [F(CX^\Lambda, 256)] = F_{\text{top}} + \bar{F}_{\text{top}} \quad \text{and} \\ Z_{BH}(\phi^\Lambda, p^\Lambda) &= |Z_{\text{top}}(t^A, g_{\text{top}})|^2, \quad \text{with} \\ t^A &= \frac{p^A + \mathbf{i}\phi^A/\pi}{p^0 + \mathbf{i}\phi^0/\pi}, \quad g_{\text{top}} = \pm \frac{4\pi \mathbf{i}}{p^0 + \mathbf{i}\phi^0/\pi}. \end{aligned}$$

Supergravity Approach to Z_{BH} . The above relation

$$Z_{BH} = |Z_{\text{top}}|^2 \quad (6.343)$$

can have a simpler super-gravity derivation [Ooguri *et. al.* (2004)].

A main ingredient in this derivation is the observation that the $N = 2$ super-gravity coupled to vector multiplets can be written as the action

$$S = \int d^4x d^4\theta (\text{super} - \text{volume form}) + h.c. = \int d^4x \sqrt{-g} R + \dots, \quad (6.344)$$

where the super-volume form in the above depends non-trivially on curvature of the fields. This reproduces the ordinary action after integrating over $d^4\theta$ and picking up the θ^4 term in the super-volume. In the context of black holes the boundary terms accompanying (6.344) give the classical black hole entropy.

We now become the derivation of (6.343). As was observed in [Bershadsky *et. al.* (1994); Antoniadis *et. al.* (1994)], topological string computes the terms

$$F = \sum_{h=0}^{\infty} \int d^4x d^4\theta F_h(X) (W^2)^g + c.c. \quad (6.345)$$

There are various terms one can get from the above action after integrating over $d^4\theta$. Let us concentrate on one of the terms which turns out to be the relevant one for us: Take the top components of X^Λ and W^2 , and absorb the $d^4\theta$ integral from the super-volume measure as in (6.344). We will work in the gauge $X^0 \sim 1$ and thus $C \sim 1/g_{\text{top}}$. As noted before in the near-horizon black hole geometry in this gauge the top component $W^2 \sim 1/C^2 \sim g_{\text{top}}^2$ and the X^Λ are fixed by the attractor mechanism. Thus, we have the black hole free energy

$$\begin{aligned} \ln Z_{BH} &= \sum_{h=0}^{\infty} g_{\text{top}}^{2h} F_{\text{top},h}(X^\Lambda/X^0) \int d^4x d^4\theta + c.c. \\ &= \sum_{g=0}^{\infty} (g_{\text{top}})^{2h-2} F_{\text{top},h}(X^\Lambda/X^0) + c.c. \\ &= 2 \operatorname{Re} F_{\text{top}}, \quad (\text{using } \int d^4x d^4\theta \sim 1/g_{\text{top}}^2). \end{aligned}$$

Upon exponentiation this leads to (6.343).

Here we have shown that if we consider one absorption of θ^4 term in (6.345) upon $d^4\theta$ integral we get the desired result. That there be no other

terms is not obvious. For example another way to absorb the θ 's would have given the familiar term $R^2 F^{2g-2}$ where F is the graviphoton field. However, such terms do not contribute in the black hole background. It would be nice to find a simple way to argue why these terms do not contribute and that we are left with this simple absorption of the θ integrals.

6.8 Application: Advanced Geometry and Topology of String Theory

6.8.1 String Theory and Noncommutative Geometry

The idea that the space–time coordinates do not commute is quite old (see [Snyder (1947a); Snyder (1947b)]). It has been studied by many authors both from a mathematical and a physical perspective. The theory of operator algebras has been suggested as a framework for physics in noncommutative space–time (see [Connes (1994)] for an exposition of the philosophy). Yang–Mills (YM) theory on a noncommutative torus has been proposed as an example; though this example at first sight appears to be neither covariant nor causal, it has proved to arise in string theory in a definite limit [Connes *et. al.* (1997)] with the non–covariance arising from the expectation value of a background field. This analysis involved toroidal compactification, in the limit of small volume, with fixed and generic values of the world–sheet theta angles. This limit is fairly natural in the context of the matrix model of M –theory [Matacz (2002)], and the original discussion was made in this context. Part of the beauty of this analysis in [Connes *et. al.* (1997)] was that T –duality acts within the noncommutative YM framework, rather than mixing the modes of noncommutative YM theory with string winding states and other stringy excitations. This makes the framework of noncommutative YM theory seem very powerful.

Seiberg and Witten in [Seiberg and Witten (1999)], reexamined the quantization of open strings ending on D –branes in the presence of a B –field. They have showed that noncommutative YM theory is valid for some purposes in the presence of any nonzero constant B –field, and that there is a systematic and efficient description of the physics in terms of noncommutative YM theory when B is large. The limit of a torus of small volume with fixed theta angle (that is, fixed periods of B) is an example with large B , but it is also possible to have large B on \mathbb{R}^n and thereby make contact with the application of noncommutative YM to instantons on \mathbb{R}^4 . An important element in their analysis is a distinction between two

different metrics in the problem. Distances measured with respect to one metric have been scaled to zero. However, the noncommutative theory is on a space with a different metric with respect to which all distances are nonzero. This guarantees that both on \mathbb{R}^n and on T^n we end up with a theory with finite metric.

6.8.1.1 Noncommutative Gauge Theory

For \mathbb{R}^n with coordinates x^i whose commutators are c -numbers, we write

$$[x^i, x^j] = i\theta^{ij},$$

with real θ . Given such a Lie algebra, one seeks to deform the algebra of functions on \mathbb{R}^n to a noncommutative, associative algebra \mathcal{A} such that

$$f * g = fg + \frac{1}{2}i\theta^{ij}\partial_i f \partial_j g + \mathcal{O}(\theta^2),$$

with the coefficient of each power of θ being a local differential expression bilinear in f and g . The essentially unique solution of this problem (modulo redefinitions of f and g that are local order by order in θ) is given by the explicit formula

$$f(x) * g(x) = e^{\frac{i}{2}\theta^{ij}\frac{\partial}{\partial x^i}\frac{\partial}{\partial x^j}} f(x+\xi)g(x+\zeta)|_{\xi=\zeta=0} = fg + \frac{1}{2}i\theta^{ij}\partial_i f \partial_j g + \mathcal{O}(\theta^2).$$

This formula defines what is often called the *Moyal bracket* of functions; it has appeared in the physics literature in many contexts, including applications to old and new matrix theories [Witten (1986b); Seiberg and Witten (1999)]. We also consider the case of $N \times N$ matrix-valued functions f, g . In this case, we define the $*$ product to be the tensor product of matrix multiplication with the $*$ product of functions as just defined. The extended $*$ product is still associative.

The $*$ product is compatible with integration in the sense that for functions f, g that vanish rapidly enough at infinity, so that one can integrate by parts in evaluating the following integrals, one has

$$\int \text{Tr } f * g = \int \text{Tr } g * f,$$

where Tr is the ordinary trace of the $N \times N$ matrices, and \int is the ordinary integration of functions.

Recall that for ordinary YM-theory, we can write the *gauge transformations* and *field strength* as

$$\delta_\lambda A_i = \partial_i \lambda + i[\lambda, A_i], \quad F_{ij} = \partial_i A_j - \partial_j A_i - i[A_i, A_j], \quad \delta_\lambda F_{ij} = i[\lambda, F_{ij}], \quad (6.346)$$

where A and λ are $N \times N$ hermitian matrices. The *Wilson line* is

$$W(a, b) = P e^{i \int_b^a A},$$

where in the path ordering $A(b)$ is to the right. Under the gauge transformation (6.346), we have

$$\delta W(a, b) = i\lambda(a)W(a, b) - iW(a, b)\lambda(b).$$

For noncommutative gauge theory, one uses the same formulas for the gauge transformation law and the field strength, except that matrix multiplication is replaced by the $*$ product. Thus, the gauge parameter $\hat{\lambda}$ takes values in \mathcal{A} tensored with $N \times N$ hermitian matrices, for some N , and the same is true for the components \hat{A}_i of the gauge field \hat{A} . The gauge transformations and field strength of *noncommutative Yang-Mills theory* are thus

$$\begin{aligned} \delta_{\hat{\lambda}} \hat{A}_i &= \partial_i \hat{\lambda} + i\hat{\lambda} * \hat{A}_i - i\hat{A}_i * \hat{\lambda}, \\ \hat{F}_{ij} &= \partial_i \hat{A}_j - \partial_j \hat{A}_i - i\hat{A}_i * \hat{A}_j + i\hat{A}_j * \hat{A}_i, \\ \delta_{\hat{\lambda}} \hat{F}_{ij} &= i\hat{\lambda} * \hat{F}_{ij} - i\hat{F}_{ij} * \hat{\lambda}. \end{aligned}$$

The theory obtained this way reduces to conventional $U(N)$ YM theory for $\theta \rightarrow 0$. Because of the way that the theory is constructed from associative algebras, there seems to be no convenient way to get other gauge groups. The commutator of two infinitesimal gauge transformations with generators $\hat{\lambda}_1$ and $\hat{\lambda}_2$ is, rather as in ordinary YM theory, a gauge transformation generated by $i(\hat{\lambda}_1 * \hat{\lambda}_2 - \hat{\lambda}_2 * \hat{\lambda}_1)$. Such commutators are nontrivial even for the rank 1 case, that is $N = 1$, though for $\theta = 0$ the rank 1 case is the Abelian $U(1)$ gauge theory. For rank 1, to first order in θ , the above formulas for the gauge transformations and field strength read

$$\begin{aligned} \delta_{\hat{\lambda}} \hat{A}_i &= \partial_i \hat{\lambda} - \theta^{kl} \partial_k \hat{\lambda} \partial_l \hat{A}_i + \mathcal{O}(\theta^2), \\ \hat{F}_{ij} &= \partial_i \hat{A}_j - \partial_j \hat{A}_i + \theta^{kl} \partial_k \hat{A}_i \partial_l \hat{A}_j + \mathcal{O}(\theta^2), \\ \delta_{\hat{\lambda}} \hat{F}_{ij} &= -\theta^{kl} \partial_k \hat{\lambda} \partial_l \hat{F}_{ij} + \mathcal{O}(\theta^2). \end{aligned}$$

6.8.1.2 Open Strings in the Presence of Constant B -Field

Bosonic Strings

Following [Seiberg and Witten (1999)], we now study strings in flat space, with metric g_{ij} , in the presence of a constant *Neveu–Schwarz B -field* and with Dp -branes. The B -field is equivalent to a constant magnetic field on the brane.

We denote the rank of the matrix B_{ij} as r ; r is of course even. Since the components of B not along the brane can be gauged away, we can assume that $r \leq p + 1$. When our target space has Lorentzian signature, we will assume that $B_{0i} = 0$, with “0” the time direction. With a Euclidean target space we will not impose such a restriction. Our discussion applies equally well if space is \mathbb{R}^{10} or if some directions are toroidally compactified with $x^i \sim x^i + 2\pi r^i$. (One could pick a coordinate system with $g_{ij} = \delta_{ij}$, in which case the identification of the compactified coordinates may not be simply $x^i \sim x^i + 2\pi r^i$, but we will not do that.) If our space is \mathbb{R}^{10} , we can pick coordinates so that B_{ij} is nonzero only for $i, j = 1, \dots, r$ and that g_{ij} vanishes for $i = 1, \dots, r$, $j \neq 1, \dots, r$. If some of the coordinates are on a torus, we cannot pick such coordinates without affecting the identification $x^i \sim x^i + 2\pi r^i$. For simplicity, we will still consider the case $B_{ij} \neq 0$ only for $i, j = 1, \dots, r$ and $g_{ij} = 0$ for $i = 1, \dots, r$, $j \neq 1, \dots, r$.

The world-sheet action is

$$\begin{aligned} S &= \frac{1}{4\pi\alpha'} \int_{\Sigma} (g_{ij} \partial_a x^i \partial^a x^j - 2\pi i \alpha' B_{ij} \epsilon^{ab} \partial_a x^i \partial_b x^j) \\ &= \frac{1}{4\pi\alpha'} \int_{\Sigma} g_{ij} \partial_a x^i \partial^a x^j - \frac{i}{2} \int_{\partial\Sigma} B_{ij} x^i \partial_t x^j, \end{aligned} \quad (6.347)$$

where Σ is the string world-sheet, which we take to be with Euclidean signature. ∂_t is a tangential derivative along the world-sheet boundary $\partial\Sigma$. The equations of motion determine the boundary conditions. For i along the Dp -branes they are

$$g_{ij} \partial_n x^j + 2\pi i \alpha' B_{ij} \partial_t x^j|_{\partial\Sigma} = 0, \quad (6.348)$$

where ∂_n is a normal derivative to $\partial\Sigma$. These boundary conditions are not compatible with real x , though with a Lorentzian world-sheet the analogous boundary conditions would be real. Nonetheless, the open string theory can be analyzed by determining the propagator and computing the correlation functions with these boundary conditions. In fact, another approach to the

open string problem is to omit or not specify the boundary term with B in the action (6.347) and simply impose the boundary conditions (6.348).

For $B = 0$, the boundary conditions in (6.348) are Neumann boundary conditions. When B has rank $r = p$ and $B \rightarrow \infty$, or equivalently $g_{ij} \rightarrow 0$ along the spatial directions of the brane, the boundary conditions become Dirichlet; indeed, in this limit, the second term in (6.348) dominates, and, with B being invertible, (6.348) reduces to $\partial_t x^j = 0$. This interpolation from Neumann to Dirichlet boundary conditions will be important, since we will eventually take $B \rightarrow \infty$ or $g_{ij} \rightarrow 0$. For B very large or g very small, each boundary of the string world-sheet is attached to a single point in the Dp -brane, as if the string is attached to a zero-brane in the Dp -brane. Intuitively, these zero-branes are roughly the constituent zero-branes of the Dp -brane as in the matrix model of M -theory [Seiberg and Witten (1999)], an interpretation that is supported by the fact that in the matrix model the construction of Dp -branes requires a nonzero B -field.

Our main focus is the case that Σ is a disc, corresponding to the classical approximation to open string theory. The disc can be conformally mapped to the upper half plane; in this description, the boundary conditions (6.348) are

$$g_{ij}(\partial - \bar{\partial})x^j + 2\pi\alpha' B_{ij}(\partial + \bar{\partial})x^j|_{z=\bar{z}} = 0,$$

where $\partial = \partial/\partial z$, $\bar{\partial} = \partial/\partial \bar{z}$, and $\text{Im } z \geq 0$. The propagator with these boundary conditions is

$$\begin{aligned} \langle x^i(z)x^j(z') \rangle = & -\alpha' [g^{ij} \log |z - z'| - g^{ij} \log |z - \bar{z}'|] \\ & + G^{ij} \log |z - \bar{z}'|^2 + \frac{1}{2\pi\alpha'} \theta^{ij} \log \frac{z - \bar{z}'}{\bar{z} - z'} + D^{ij}]. \end{aligned} \quad (6.349)$$

Here

$$\begin{aligned} G^{ij} &= \left(\frac{1}{g + 2\pi\alpha' B} \right)_S^{ij} = \left(\frac{1}{g + 2\pi\alpha' B} g \frac{1}{g - 2\pi\alpha' B} \right)^{ij}, \\ G_{ij} &= g_{ij} - (2\pi\alpha')^2 (B g^{-1} B)_{ij}, \\ \theta^{ij} &= 2\pi\alpha' \left(\frac{1}{g + 2\pi\alpha' B} \right)_A^{ij} = -(2\pi\alpha')^2 \left(\frac{1}{g + 2\pi\alpha' B} B \frac{1}{g - 2\pi\alpha' B} \right)^{ij}, \end{aligned}$$

where $(\)_S$ and $(\)_A$ denote the symmetric and antisymmetric part of the matrix. The constants D^{ij} in (6.349) can depend on B but are independent of z and z' ; they play no essential role and can be set to a convenient value. The first three terms in (6.349) are manifestly single-valued. The fourth

term is single-valued, if the branch cut of the logarithm is in the lower half plane.

Our focus is on the open string vertex operators and interactions [Witten (1986b); Seiberg and Witten (1999)]. Open string vertex operators are of course inserted on the boundary of Σ . So to get the relevant propagator, we restrict (6.349) to real z and z' , which we denote τ and τ' . Evaluated at boundary points, the propagator is

$$\langle x^i(\tau)x^j(\tau') \rangle = -\alpha' G^{ij} \log(\tau - \tau')^2 + \frac{i}{2} \theta^{ij} \epsilon(\tau - \tau'), \quad (6.350)$$

where we have set D^{ij} to a convenient value. $\epsilon(\tau)$ is the function that is 1 or -1 for positive or negative τ .

The object G_{ij} has a very simple intuitive interpretation: it is the effective metric seen by the open strings. The short distance behavior of the propagator between interior points on Σ is

$$\langle x^i(z)x^j(z') \rangle = -\alpha' g^{ij} \log |z - z'|.$$

The coefficient of the logarithm determines the anomalous dimensions of closed string vertex operators, so that it appears in the mass shell condition for closed string states. Thus, we will refer to g_{ij} as the closed string metric. G_{ij} plays exactly the analogous role for open strings, since anomalous dimensions of open string vertex operators are determined by the coefficient of $\log(\tau - \tau')^2$ in (6.350), and in this coefficient G^{ij} enters in exactly the way that g^{ij} would enter at $\theta = 0$. We will refer to G_{ij} as the open string metric.

The coefficient θ^{ij} in the propagator also has a simple intuitive interpretation. In conformal field theory, one can compute commutators of operators from the short distance behavior of operator products by interpreting time ordering as operator ordering. Interpreting τ as time, we see that

$$[x^i(\tau), x^j(\tau')] = T(x^i(\tau)x^j(\tau^-) - x^i(\tau)x^j(\tau^+)) = i\theta^{ij}. \quad (6.351)$$

That is, x^i are coordinates on a noncommutative space with noncommutativity parameter θ .

Consider the product of tachyon vertex operators $e^{ip \cdot x}(\tau)$ and $e^{iq \cdot x}(\tau')$. With $\tau > \tau'$, we get for the leading short distance singularity

$$e^{ip \cdot x}(\tau) \cdot e^{iq \cdot x}(\tau') \sim (\tau - \tau')^{2\alpha' G^{ij} p_i q_j} e^{-\frac{1}{2} i \theta^{ij} p_i q_j} e^{i(p+q) \cdot x}(\tau') + \dots$$

If we could ignore the term $(\tau - \tau')^{2\alpha' p \cdot q}$, then the formula for the operator

product would reduce to a $*$ product; we would get

$$e^{ip \cdot x}(\tau) e^{iq \cdot x}(\tau') \sim e^{ip \cdot x} * e^{iq \cdot x}(\tau').$$

This is no coincidence. If the dimensions of all operators were zero, the leading terms of operator products $\mathcal{O}(\tau)\mathcal{O}'(\tau')$ would be independent of $\tau - \tau'$ for $\tau \rightarrow \tau'$, and would give an ordinary associative product of multiplication of operators. This would have to be the $*$ product, since that product is determined by associativity, translation invariance, and (6.351) (in the form $x^i * x^j - x^j * x^i = i\theta^{ij}$).

Now, consider an operator on the boundary of the disc that is of the general form $P(\partial x, \partial^2 x, \dots) e^{ip \cdot x}$, where P is a polynomial in derivatives of x , and x are coordinates along the Dp -brane (the transverse coordinates satisfy Dirichlet boundary conditions). Since the second term in the propagator (6.350) is proportional to $\epsilon(\tau - \tau')$, it does not contribute to contractions of derivatives of x . Therefore, the expectation value of a product of k such operators, of momenta p^1, \dots, p^k , satisfies

$$\left\langle \prod_{n=1}^k P_n(\partial x(\tau_n), \partial^2 x(\tau_n), \dots) e^{ip^n \cdot x(\tau_n)} \right\rangle_{G, \theta} = \quad (6.352)$$

$$e^{-\frac{i}{2} \sum_{n>m} p_i^n \theta^{ij} p_j^m \epsilon(\tau_n - \tau_m)} \left\langle \prod_{n=1}^k P_n(\partial x(\tau_n), \partial^2 x(\tau_n), \dots) e^{ip^n \cdot x(\tau_n)} \right\rangle_{G, \theta=0},$$

where $\langle \dots \rangle_{G, \theta}$ is the expectation value with the propagator (6.350) parametrized by G and θ . We see that when the theory is described in terms of the open string parameters G and θ , rather than in terms of g and B , the θ dependence of correlation functions is very simple. Note that because of momentum conservation ($\sum_m p^m = 0$), the crucial factor

$$\exp \left(-\frac{i}{2} \sum_{n>m} p_i^n \theta^{ij} p_j^m \epsilon(\tau_n - \tau_m) \right) \quad (6.353)$$

depends only on the cyclic ordering of the points τ_1, \dots, τ_k around the circle [Witten (1986b); Seiberg and Witten (1999)].

The string theory S -matrix can be obtained from the conformal field theory correlators by putting external fields on shell and integrating over the τ 's. Therefore, it has a structure inherited from (6.352). To be very precise, in a theory with $N \times N$ Chan-Paton factors, consider a k point function of particles with Chan-Paton wave functions W_i , $i = 1, \dots, k$, momenta p_i , and additional labels such as polarizations or spins that we will

generically call ϵ_i . The contribution to the scattering amplitude in which the particles are cyclically ordered around the disc in the order from 1 to k depends on the Chan–Paton wave functions by a factor $\text{Tr}(W_1 W_2 \dots W_k)$. We suppose, for simplicity, that N is large enough so that there are no identities between this factor and similar factors with other orderings. By studying the behavior of the S –matrix of massless particles of small momenta, one can extract order by order in α' a low energy effective action for the theory. If Φ_i is an $N \times N$ matrix-valued function in space–time representing a wavefunction for the i^{th} field, then at $B = 0$ a general term in the effective action is a sum of expressions of the form

$$\int d^{p+1}x \sqrt{\det G} \text{Tr}(\partial^{n_1} \Phi_1 \partial^{n_2} \Phi_2 \dots \partial^{n_k} \Phi_k). \quad (6.354)$$

Here ∂^{n_i} is, for each i , the product of n_i partial derivatives with respect to some of the space–time coordinates; which coordinates it is has not been specified in the notation. The indices on fields and derivatives are contracted with the metric G , though this is not shown explicitly in the formula.

Now to incorporate the B –field, at fixed G , is very simple: if the effective action is written in momentum space, we need only incorporate the factor (6.353). Including this factor is equivalent to replacing the ordinary product of fields in (6.354) by a $*$ product (in this formulation, one can work in coordinate space rather than momentum space). So the term corresponding to (6.354) in the effective action is given by the same expression but with the wave functions multiplied using the $*$ product:

$$\int d^{p+1}x \sqrt{\det G} \text{Tr}(\partial^{n_1} \Phi_1 * \partial^{n_2} \Phi_2 * \dots * \partial^{n_k} \Phi_k).$$

It follows, then, that the B dependence of the effective action for fixed G and constant B can be obtained in the following very simple fashion: replace ordinary multiplication by the $*$ product. We will make presently an explicit calculation of an S –matrix element to illustrate this statement, and we will make a detailed check of a different kind in section 4 using almost constant fields and the *Dirac–Born–Infeld theory*.

Now, recall that background gauge fields couple to the string world–sheet by adding

$$-i \int d\tau A_i(x) \partial_\tau x^i \quad (6.355)$$

to the action (6.347). We assume for simplicity that there is only a rank one gauge field; the extension to higher rank is straightforward. Comparing (6.347) and (6.355), we see that a constant B -field can be replaced by the gauge field $A_i = -\frac{1}{2}B_{ij}x^j$, whose field strength is $F = B$. When we are working on \mathbb{R}^n , we are usually interested in situations where B and F are constant at infinity, and we fix the ambiguity by requiring that F is zero at infinity.

Naively, (6.355) is invariant under ordinary gauge transformations as

$$\delta A_i = \partial_i \lambda, \quad (6.356)$$

because (6.355) transforms by a total derivative

$$\delta \int d\tau A_i(x) \partial_\tau x^i = \int d\tau \partial_i \lambda \partial_\tau x^i = \int d\tau \partial_\tau \lambda.$$

However, because of the infinities in quantum field theory, the theory has to be regularized and we need to be more careful. We will examine a point splitting regularization, where different operators are never at the same point.

Then expanding the exponential of the action in powers of A and using the transformation law (6.356), we find that the functional integral transforms by

$$- \int d\tau A_i(x) \partial_\tau x^i \cdot \int d\tau' \partial_{\tau'} \lambda, \quad (6.357)$$

plus terms of higher order in A . The product of operators in (6.357) can be regularized in a variety of ways. We will make a point-splitting regularization in which we cut out the region $|\tau - \tau'| < \delta$ and take the limit $\delta \rightarrow 0$. Though the integrand is a total derivative, the τ' integral contributes surface terms at $\tau - \tau' = \pm\delta$. In the limit $\delta \rightarrow 0$, the surface terms contribute

$$\begin{aligned} & - \int d\tau A_i(x(\tau)) \partial_\tau x^i(\tau) (\lambda(x(\tau^-)) - \lambda(x(\tau^+))) \\ & = - \int d\tau (A_i(x) * \lambda - \lambda * A_i(x)) \partial_\tau x^i. \end{aligned}$$

Here we have used the relation of the operator product to the $*$ product, and the fact that with the limit boundary propagator [Seiberg and Witten (1999)]

$$\langle x^i(\tau) x^j(0) \rangle = \frac{i}{2} \theta^{ij} \epsilon(\tau),$$

there is no contraction between $\partial_\tau x$ and x . To cancel this term, we must add another term to the variation of the gauge field; the theory is invariant not under (6.356), but under

$$\delta \hat{A}_i = \partial_i \lambda + i\lambda * \hat{A}_i - i\hat{A}_i * \lambda. \quad (6.358)$$

This is the gauge invariance of noncommutative YM theory, and in recognition of that fact we henceforth denote the gauge field in the theory defined with point splitting regularization as \hat{A} . A sigma model expansion with Pauli–Villars regularization would have preserved the standard gauge invariance of open string gauge field, so whether we get ordinary or noncommutative gauge fields depends on the choice of regulator.

We have made this derivation to lowest order in \hat{A} , but it is straightforward to go to higher orders. At the n th order in \hat{A} , the variation is

$$\begin{aligned} \frac{i^{n+1}}{n!} \int \hat{A}(x(t_1)) \dots \hat{A}(x(t_n)) \partial_t \lambda(x(t)) \\ + \frac{i^{n+1}}{(n-1)!} \int \hat{A}(x(t_1)) \dots \hat{A}(x(t_{n-1})) \left(\lambda * \hat{A}(x(t_n)) - \hat{A} * \lambda(x(t_n)) \right), \end{aligned} \quad (6.359)$$

where the integration region excludes points where some t 's coincide. The first term in (6.359) arises by using the naive gauge transformation (6.356), and expanding the action to n th order in \hat{A} and to first order in λ . The second term arises from using the correction to the gauge transformation in (6.358) and expanding the action to the same order in \hat{A} and λ . The first term can be written as

$$\begin{aligned} \frac{i^{n+1}}{n!} \sum_j \int \hat{A}(x(t_1)) \dots \hat{A}(x(t_{j-1})) \hat{A}(x(t_{j+1})) \dots \\ \dots \hat{A}(x(t_n)) \left(\hat{A} * \lambda(x(t_j)) - \lambda * \hat{A}(x(t_j)) \right) \\ = \frac{i^{n+1}}{(n-1)!} \int \hat{A}(x(t_1)) \dots \hat{A}(x(t_{n-1})) \left(\hat{A} * \lambda(x(t_n)) - \lambda * \hat{A}(x(t_n)) \right), \end{aligned}$$

making it clear that (6.359) vanishes. Therefore, there is no need to modify the gauge transformation law (6.358) at higher orders in \hat{A} . For further technical details, see [Witten (1986b); Seiberg and Witten (1999)].

6.8.2 K –Theory Classification of Strings

In this subsection, following [Witten (1998c); Witten (2000)], we will revisit the relation between strings, D –brane and K –theory, which we started in

section 4.5.

As Ed. Witten has pointed out, K -theory provides a framework for classifying Ramond–Ramond (RR) charges and fields. K -theory of manifolds has a natural extension to K -theory of noncommutative algebras, such as the algebras considered in noncommutative Yang–Mills (YM) theory, or in open string field theory. In a number of concrete problems, the K -theory analysis proceeds most naturally if one starts out with an infinite set of D -branes, reduced by tachyon condensation to a finite set.

Recall that a D -brane wrapped on a submanifold S of space–time may carry a nonzero RR-charge. As RR-fields are p -forms, superficially it seems that the conserved charge should be measured by the cohomology class of the RR-form (or of the cycle S itself). However, D -branes carry gauge fields; and gauge fields are not natural in ordinary (co)homology theories. Instead, they are natural in K -theory.

More precisely, if X is space–time and $\mathcal{A}(X)$ is the commutative, associative algebra of continuous complex-valued functions on X , then the K -theory of X can be defined in terms of representations of $\mathcal{A}(X)$. A representation of a ring is usually called a *module*. Here are some examples of $\mathcal{A}(X)$ -modules.

The most obvious example of an $\mathcal{A}(X)$ -module is $\mathcal{A}(X)$ itself. For $f \in \mathcal{A}(X)$ (regarded as a ring) and $g \in \mathcal{A}(X)$ (regarded as a module), we define $f(g) = fg$, where on the right hand side the multiplication occurs in $\mathcal{A}(X)$. This obviously obeys the defining condition of a module, which is that

$$(f_1 f_2)(g) = f_1(f_2(g)).$$

More generally, consider a Dp -brane (or a collection of N Dp -branes for some $N > 0$) wrapped on a submanifold S of X , with any *Chan–Paton gauge bundle* W on the D -brane. Let $M(S)$ be the space of sections of W , that is, the space of one-particle states for a charged scalar coupled to the bundle W . Then $M(S)$ is an $\mathcal{A}(X)$ -module; for $f \in \mathcal{A}(X)$, $g \in M(S)$, we simply set again $f(g) = fg$. On the right hand side, the multiplication is defined by restricting f , which is a function on X , to S and then multiplying f and g .

So in, say, Type IIB superstring theory, a collection of $D9$ -branes defines a representation or module E of $\mathcal{A}(X)$. A collection of $\overline{D9}$ -branes defines another module F . So any configuration of $D9$ and $\overline{D9}$ -branes determines a pair (E, F) .

To classify D -brane charge, we want to classify pairs (E, F) modulo physical processes. An important process [Sen (1998)] is brane–antibrane

creation and annihilation – the creation or annihilation of a set of $D9$'s and $\overline{D9}$'s each bearing the same gauge bundle G . This amounts to

$$(E, F) \leftrightarrow (E \oplus G, F \oplus G). \quad (6.360)$$

The equivalence classes make up a group called $K(X)$ (or $K(\mathcal{A}(X))$ if we want to make the interpretation in terms of $\mathcal{A}(X)$ -modules more explicit). The addition law in this group is just

$$(E, F) + (E' F') = (E \oplus E', F \oplus F').$$

The inverse of (E, F) is (F, E) ; note that $(E, F) \oplus (F, E) = (E \oplus F, E \oplus F)$, and using the equivalence relation (6.360), this is equivalent to zero. D -branes of Type IIB carry conserved charges that take values in $K(X)$ [Witten (1998c)]. In the above definition of $K(X)$, we used only ninebranes, even though, as we explained earlier, an $\mathcal{A}(X)$ module can be constructed using Dp -branes (or antibranes) for any p . In fact, we can classify D -brane charge just using the ninebranes, and then build the Dp -branes of $p < 9$ via pairs (E, F) with a suitable tachyon condensate.

Wherever one looks closely at topological properties of RR charges (or fields), one sees effects that reflect the K -theory structure. For example, there are stable D -brane states (like the nonsupersymmetric $D0$ -branes of Type I) that would not exist if D -brane charge were classified by cohomology instead of K -theory. Conversely, it is possible to have a D -brane state that would be stable if D -brane charge were measured by cohomology, but which is in fact unstable (via a process that involves nucleation of $D9 - \overline{D9}$ pairs in an intermediate state). This occurs in Type II superstring theory, in which a D -brane wrapped on a homologically nontrivial cycle in space-time is in fact in certain cases unstable.

According to Witten, there is a deeper reason that it is good to know about the K -theory interpretation of D -branes: it may be naturally adapted for string generalizations. In fact, we can define $K(X)$ in terms of representations of the algebra $\mathcal{A}(X)$ of functions on space-time. We can similarly define $K(\mathcal{A})$ for any noncommutative algebra \mathcal{A} , in terms of pairs (E, F) of \mathcal{A} -modules. By contrast, we would not have an equally useful and convenient notion of 'cohomology' if the algebra of functions on space-time is replaced by a noncommutative ring.

For example, turning on a *Neveu-Schwarz B-field*, we can make $\mathcal{A}(X)$ noncommutative; the associated $K(\mathcal{A})$ was used by Connes, Douglas, and Schwarz in the original paper on noncommutative YM-theory applied to

string theory [Connes *et. al.* (1997)]. This is an interesting example, even though it involves only the zero modes of the strings. One would much like to have a fully stringy version involving a noncommutative algebra constructed using all of the modes of the string, not just the zero modes.

At this stage, we do not know what is the right noncommutative algebra that uses all of the modes. One concrete candidate is the $*$ -algebra of open string field theory, defined in terms of gluing strings together. If we call this algebra \mathcal{A}_{st} , it seems plausible that D -brane charge is naturally labelled by $K(\mathcal{A}_{st})$. For a manifold of very large volume compared to the string scale, we would conjecture that $K(\mathcal{A}_{st})$ is the same as the ordinary $K(X)$ of topological K -theory.

K -Theory and RR-Fields

Naively speaking, an RR p -form field G_p obeys a Dirac quantization law according to which, for any p -cycle U in space-time,

$$\int_U \frac{G_p}{2\pi} = \text{integer}. \quad (6.361)$$

If that were the right condition, then RR fields would be classified by cohomology. However, that is not the right answer, because the actual quantization condition on RR periods is much more subtle than (6.361). There are a variety of corrections to (6.361) that involve space-time curvature and the gauge fields on the brane, as well as self-duality and global anomalies.

The answer, for Type IIB superstrings, turns out to be that RR fields are classified by $K^1(X)$. For our purposes, $K^1(X)$ can be defined as the group of components of the group of continuous maps from X to $U(N)$, for any sufficiently large N . This means that topological classes of RR fields on X are classified by a map $U : X \rightarrow U(N)$ for some large N . The relation of G_p to U is roughly $G_p \sim \text{Tr}(U^{-1}dU)^p$, where we have ignored corrections due to space-time curvature and subtleties associated with self-duality of RR fields [Fabinger and Horava (2000)].

The physical meaning of U is not clear. For Type IIA, the analog is that RR fields are classified by a $U(N)$ gauge bundle (for some large N) with connection A and curvature F_A , the relation being $G_p \sim \text{Tr } F_A^{p/2}$. The analog for M -theory involves E_8 gauge bundles with connection. Again, the physical meaning of the $U(N)$ or E_8 gauge fields is not clear.

The value of using K^1 to classify RR fields of Type IIB is that this gives a concise way to summarize the otherwise rather complicated quantization

conditions obeyed by the RR fields. In addition, this framework is useful in describing subtle phase factors that enter in the RR partition function. In hindsight, once it is known that RR charges are classified by K -theory, one should have suspected a similar classification for RR fields. After all, RR charges produce RR fields! So the geometry used to classify RR charges must be similar to the geometry used to classify RR fields.

Just like $K(X)$, $K^1(X)$ has an analog for any noncommutative algebra \mathcal{A} . Given \mathcal{A} , we let \mathcal{A}_N denote the group of invertible $N \times N$ matrices whose matrix elements are elements of \mathcal{A} . Then $K^1(\mathcal{A})$ is the group of components of \mathcal{A}_N , for large N .

For example, for $\mathcal{A} = \mathcal{A}(X)$ the ring of complex-valued continuous functions on X , \mathcal{A}_N is the group of maps of X to $GL(N, \mathbf{C})$. This is contractible to the group of maps of X to $U(N)$, so for large N the group of components of \mathcal{A}_N is the same as $K^1(X)$, as we defined it initially.

The existence of a generalization of $K^1(X)$ for noncommutative rings means that the description of Type IIB RR fields by $K^1(X)$ in the long distance limit may be a useful starting point for stringy generalizations.

$N \rightarrow \infty$

In the previous subsection, N was a sufficiently large but finite integer. Our next task will be to describe some things that depend on setting N equal to infinity.

Before doing so, let us recall the role of the $N \rightarrow \infty$ limit in physics. It is important in the conjectured link of gauge theory with strings; in the old matrix models that are used to give soluble examples of string theory; in the matrix model of M -theory; and in the correspondence between gravity in an asymptotically AdS space-time and conformal field theory on the boundary.

For Type IIB superstrings, [Witten (1998c); Witten (2000)] used $K(X)$ to classify RR charges, and $K^1(X)$ to classify RR fields.

The T -dual statement is that for Type IIA, $K^1(X)$ should classify RR-charges, and $K(X)$ should classify RR fields. Recall that, by Bott periodicity, $K^{i+2}(X) = K^i(X)$, so the only K -groups of X are $K^0(X)$, which we have called simply $K(X)$, and $K^1(X)$.

The most concrete and natural attempt to explain in general why $K^1(X)$ classifies RR charges for Type IIA is that of [Horava (1999)]. The starting point here is to consider a system of N unstable $D9$ -branes of Type IIA. The branes support a $U(N)$ gauge field and a tachyon field T in the adjoint

representation of $U(N)$. There is a symmetry $T \rightarrow -T$.

The effective potential for the tachyon field is believed to have the general form

$$V(T) = \frac{1}{g_{st}} \operatorname{Tr} F(T),$$

where the function $F(T)$ is non-negative and, after scaling T correctly, vanishes precisely if $T = \pm 1$. Hence $V(T)$ is minimized if and only if every eigenvalue of T is ± 1 .

It was argued in [Horava (1999)] that, in flat \mathbb{R}^{10} , one can make supersymmetric Dp -branes (for even p) as solitons of T . For example, to make a $D6$ -brane, we set $N = 2$. Let \vec{x} be the three coordinates normal to the $D6$ -brane, and set

$$T = \frac{\vec{\sigma} \cdot \vec{x}}{|x|} f(|x|),$$

where $f(r) \sim r$ for small r , and $f(r) \rightarrow \infty$ for $r \rightarrow \infty$. So for $|x| \rightarrow \infty$, the eigenvalues of T are everywhere ± 1 . Near $x = 0$, there is a topological knot that we interpret as the $D6$ -brane.

In flat \mathbb{R}^{10} , one can similarly make Dp -branes for other even p . But on a general space-time, this does not work for arbitrary Dp -branes unless we set $N = \infty$. The problem is most obvious if X , or at least its spatial part, is compact. The tachyon field T , being adjoint-valued, maps X to the Lie algebra of $U(N)$; since the Lie algebra is contractible, T carries no topology. So a map from X to the Lie algebra does not represent an element of $K^1(X)$; indeed, it does not carry topological information at all. To define an element of $K^1(X)$, we need the group, not the Lie algebra; a map $U : X \rightarrow U(N)$ does the job.

As Atiyah and Singer showed long ago, we get back the right topology from the Lie algebra if we set $N = \infty$! We have to interpret $U(\infty)$ to be the unitary group \mathcal{U} of a separable Hilbert space \mathcal{H} of countably infinite dimension. We interpret the $N = \infty$ analog of the space of hermitian $N \times N$ matrices to be the space of bounded self-adjoint operators T on \mathcal{H} whose spectrum is as follows: there are infinitely many positive eigenvalues and infinitely many negative ones, and zero is not an accumulation point of the spectrum. The last condition makes T a *Fredholm operator*. Physically, T should be required to obey these conditions, since they are needed to make the energy and the $D8$ -brane charge finite. In fact, to make the energy finite, almost all the eigenvalues of T are very close to ± 1 . Anyway, with

these conditions imposed on T , it turns out that the space of T 's has the same topology as that of $U(N)$ for large N .

So we can use tachyon condensation on a system of $D9$ -branes to describe RR charges for Type IIA. But we have to start with infinitely many $D9$ -branes, which then undergo tachyon condensation down to a configuration of finite energy.

Let us now explain in more concrete terms the obstruction to making D -branes in this way for finite N , and how it vanishes for $N = \infty$. Let us go back to the example of a $D6$ -brane constructed with 2 $D9$ -branes. We took the transverse directions to be a copy of \mathbb{R}^3 , and the tachyon field to be

$$T = \frac{\vec{\sigma} \cdot \vec{x}}{|x|} f(|x|).$$

If we try to compactify the transverse directions to \mathbb{S}^3 , we run into trouble because T is not constant at infinity. The conjugacy class of T is constant at infinity – the eigenvalues of T are everywhere 1 and -1 – but T itself is not constant.

Moreover, T is not homotopic to a constant at infinity. If T were homotopic to a constant near infinity, we would deform it to be constant and then extend it over \mathbb{S}^3 . But it is not homotopic to a constant.

The basic obstruction to making T constant at infinity is the ‘magnetic charge’. Let \mathbb{S}^2 be a sphere at infinity in \mathbb{R}^3 . Over \mathbb{S}^2 , we can define a line bundle L_+ whose fiber is the +1 eigenspace of T , and a line bundle L_- whose fiber is the -1 eigenspace of T . The line bundles L_+ and L_- are topologically nontrivial – their first Chern classes are respectively 1 and -1. As long as we try to deform T preserving the fact that its eigenvalues are 1 and -1, the line bundles L_{\pm} are well-defined, and their first Chern classes are invariant. So the nontriviality of L_+ (or L_-) prevents us from making a homotopy to constant T .

Let us add some additional ‘spectator’ $D9$ -branes, and see if anything changes. Suppose there are $M = 2k$ additional branes, so that the total number of branes is $N = 2 + M = 2 + 2k$. Let the tachyon field be $T' = T \oplus U$, where T is as above and U is the sum of k copies of the matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (6.362)$$

acting on the $2k$ additional branes.

Thus T' has near infinity $k + 1$ eigenvalues +1 and $k + 1$ eigenvalues -1.

The $+1$ eigenspace of T' is a vector bundle V_+ of first Chern class 1 (since it is constructed by adding a trivial bundle to L_+), and the -1 eigenspace of T' is similarly a vector bundle V_- of first Chern class -1 . In particular, V_+ and V_- are nontrivial, so we have not gained anything by adding the spectator branes: T' is not homotopic to a constant, and cannot be extended over infinity. The nontriviality of V_+ is controlled by $\pi_1(U(k+1)) = \mathbb{Z}$, which is associated with the existence of a first Chern class.

Instead, what happens if we set $k = \infty$? To be more precise, we take the number of spectator $D9$ -branes to be countably infinite, and assume $T' = T \oplus U$, where U is the direct sum of countably many copies of the matrix in (6.362). We can still define the bundles V_+ and V_- ; their fibers are separable Hilbert spaces (that is, Hilbert spaces of countably infinite dimension). $U(k+1)$ is replaced by \mathcal{U} , the unitary group of a separable Hilbert space. Now we run into the fundamental fact (Kuiper's Theorem) that \mathcal{U} is contractible; its homotopy groups are all zero. Thus, any bundle of separable Hilbert spaces is trivial. In particular, V_+ and V_- are trivial, so T' is homotopic to a constant and can be extended over infinity.

So if the total number of unstable $D9$ -branes is $N = \infty$, we can make a $D6$ -brane localized at a point in \S^3 . More generally, in view of the result of Atiyah and Singer, we can starting at $N = \infty$ build an arbitrary class in $K^1(X)$ via tachyon condensation.

In terms of applying this result to physics, there are a few issues that we should worry about. One question is simply whether it is physically sensible to start with infinitely many branes and rely on tachyon condensation to get us down to something of finite energy.

Quite a different question is whether the answer that we have obtained by setting $N = \infty$ is the right one for physics. In the field of a $D6$ -brane that is localized at a point on \S^3 , the equation for the RR two-form field G_2 (of which the $D6$ -brane is a magnetic source) has no solution, since "the flux has nowhere to go."

It seems that the situation is that $N = \infty$ corresponds to the correct answer in classical open string theory, where the effective action comes from world-sheets with the topology of a disc. The RR fields enter as a correction of relative order g_{st} (the closed string coupling constant) coming from world-sheets with cylinder topology, and should be ignored in the classical approximation.

The classification of D -branes by brane creation and annihilation holds at $g_{st} = 0$, and leads for Type IIB to a classification of D -brane charge by $K(X)$. To get the analogous answer, namely $K^1(X)$, for Type IIA via

unstable $D9$ -branes and tachyon condensation, we need to start at $N = \infty$.

Intuitively, in the absence of tachyon condensation, $N = \infty$ should correspond to $g_{st} = 0$, since the effective expansion parameter for open strings is $g_{st}N$. If N is infinite, then prior to tachyon condensation, g_{st} must be zero, or the quantum corrections diverge. If we want g_{st} to be nonzero, we need tachyon condensation to reduce to an effective finite value of N .

A somewhat analogous problem is to consider D -branes when the Neveu-Schwarz 3-form field H is topologically nontrivial. We will carry out this discussion in Type IIB (for Type IIA, we would have to combine what follows with what we said above in the absence of the H -field).

Just as at $H = 0$, we would like to classify D -brane states by pairs (E, F) (where E is a $D9$ state and F is a $\overline{D9}$ state) subject to the usual sort of equivalence relation. But there is a problem in having a $D9$ state in the presence of an H -field.

In fact, when H is topologically non-trivial, one cannot have a single $D9$ -brane. On the $D9$ -brane, there is a $U(1)$ gauge field with field strength F . The relation $dF = H$ shows, at the level of de Rham cohomology, that H must be topologically trivial if a single $D9$ -brane is present. This conclusion actually holds precisely, not just in de Rham cohomology.

There is a special case in which there is a comparatively elementary cure for this difficulty. If H is torsion, that is if there is an integer $M > 0$ such that MH is topologically trivial, then it is possible to have a set of M $D9$ -branes whose 'gauge bundle' actually has structure group $U(M)/\mathbb{Z}_M$, rather than $U(M)$. (The obstruction to lifting the $U(M)/\mathbb{Z}_M$ bundle to a $U(M)$ bundle is determined by H .) We will call such a gauge bundle a twisted bundle. More generally, for any positive integer m , we can have $N = mM$ $D9$ -branes with the structure group of the bundle being $U(mM)/\mathbb{Z}_M$. In such a situation, D -brane charge is classified, as one would guess, by pairs (E, F) of twisted bundles (or $D9$ and $\overline{D9}$ states) subject to the usual equivalence relation. The equivalence classes make a group $K_H(X)$.

If one wishes to interpret $K_H(X)$ as the K -theory of representations of an algebra, one must pick a particular twisted bundle W and consider a D -brane state with boundary conditions determined by W . The W - W open strings transform in the adjoint representation, so the gauge parameters of the zero mode sector of the open strings are sections of $W \otimes \bar{W}$. Notice that although W is a twisted bundle (with structure group $U(M)/\mathbb{Z}_M$ rather than $U(M)$), $W \otimes \bar{W}$ is an ordinary bundle, since the center acts trivially in the adjoint representation.

The sections of $W \otimes \bar{W}$ form an algebra, defined as follows: if s^i_j and t^k_l are sections of $W \otimes \bar{W}$ – where the upper and lower indices are respectively W – and \bar{W} –valued – then their product is $(st)^i_l = \sum_k s^i_k t^k_l$. This is the algebra $\mathcal{A}_W(X)$ of all endomorphisms or linear transformations of the bundle W . The algebra of open string field theory, for W – W open strings, reduces to \mathcal{A}_W if one looks only at the zero modes of the strings. This is a sensible approximation at low energies in a limit in which X is very large compared to the string scale.

If H is zero and W is a trivial rank one complex bundle, then $\mathcal{A}_W(X)$ is our friend $\mathcal{A}(X)$. If H is zero and W is a trivial rank N complex bundle, then including W means simply that there are $N \times N$ Chan–Paton matrices everywhere. So in this case, $\mathcal{A}_W(X) = \mathcal{A}(X) \otimes M_N$, where M_N is the algebra of $N \times N$ complex-valued matrices. In general, whatever H is, W is always trivial locally, so locally $\mathcal{A}_W(X)$ is isomorphic to $\mathcal{A}(X) \otimes M_N$.

A twisted bundle is equivalent to an \mathcal{A}_W –module, and the group $K_H(X)$ of pairs (E, F) of twisted bundles (modulo the usual equivalence) coincides with $K(\mathcal{A}_W)$, the K –group of \mathcal{A}_W –modules. This assertion leads to an immediate puzzle; $K_H(X)$ as defined in terms of pairs (E, F) of twisted bundles is manifestly independent of W while $K(\mathcal{A}_W)$ appears to depend on W . Indeed, given any two distinct twisted bundles W and W' , the corresponding algebras \mathcal{A}_W and $\mathcal{A}_{W'}$ are distinct, but at the same time *Morita-equivalent*.

So far, we have only considered the case that H is torsion. A typical example, important in the AdS/CFT correspondence, is the space–time $X = \text{AdS}_5 \times \text{RP}^5$, where a torsion H –field on RP^5 is used to describe $Sp(n)$ rather than $SO(2n)$ gauge theory in the boundary CFT.

However, in most physical applications, H is not torsion. In that case, we must somehow take a large M limit of what has been said above. The right way to do this has been shown by [Bouwknegt and Mathai (2000)]. The suitable large M limit of $U(M)/\mathbb{Z}_M$ is $PU(\mathcal{H}) = U(\mathcal{H})/U(1)$. In other words, for $M = \infty$, one replaces $U(M)$ by the unitary group $U(\mathcal{H})$ of a separable Hilbert space \mathcal{H} ; and one replaces \mathbb{Z}_M by $U(1)$. This means, in particular, that when H is not torsion, one cannot have a finite set of $D9$ – or $\bar{D}9$ –branes, but one can have an infinite set, with a suitable infinite rank twisted gauge bundle E or F . Then D –brane charge is classified by the group K_H of pairs (E, F) modulo the usual equivalence relation. A detailed explanation can be found in [Bouwknegt and Mathai (2000)]. Here, the Kuiper’s Theorem – the contractibility of $\mathcal{U} = U(\mathcal{H})$ – plays an

important role.

This construction, in the $M = \infty$ limit, has the beautiful property, explained in [Bouwknegt and Mathai (2000)], that the noncommutative algebra whose K -group is K_H is *unique*, independent of any arbitrary choice of twisted bundle W or W' . This really depends on the number of $D9$ and $\overline{D9}$ -branes being infinite.

This page intentionally left blank

Bibliography

- Ablowitz, M.J., Clarkson, P.A. (1991). Solitons, nonlinear evolution equations and inverse scattering. London Math. Soc., **149**, CUP, Cambridge.
- Abraham, R., Marsden, J. (1978). Foundations of Mechanics. Benjamin, Reading.
- Abraham, R., Marsden, J., Ratiu, T. (1988). Manifolds, Tensor Analysis and Applications. Springer, New York.
- Adams, J.F. (1962). Vector fields on spheres. Ann. Math. **75**, 603-632.
- Adams, J.F. (1978). Infinite Loop Spaces. Princeton Univ. Press, Princeton, NJ.
- Aidman, E.V., Leontiev, D.A. (1991). From being motivated to motivating oneself: a Vygotskian perspective. Stud. Sov. Thought, **42**, 137-151.
- Allday, C. (1998). Examples of circle actions on symplectic spaces. In Homotopy and Geometry Banach Center Publications vol 45, Institute of Mathematics, Polish Academy of Sciences, Warsaw, 87-90.
- Akbulut, S., McCarthy, J.D. (1990). Casson's invariant for homological 3-spheres-An exposition, Mathematical Notes **36**, Princeton Univ. Press, Princeton.
- Albeverio, S., Fenstat, J., Hoegh-Krohn, R., Lindstrom, T. (1986). Nonstandard Methods in Stochastic Analysis and Mathematical Physics. Academic Press, New York.
- Alekseevsky, D.V., Grabowski, J., Marmo, G., Michor, P.W. (1994). Poisson structures on the cotangent bundle of a Lie group or a principle bundle and their reductions. J. Math. Phys., **35**, 4909-4928.
- Alekseevsky, D., Grabowski, J., Marmo, G., Michor, P.W. (1997). Completely integrable systems: a generalization. Mod. Phys. Let. A, **12**(22), 1637-1648.
- Alfinito, E., Vitiello, G., (2000). Formation and life-time of memory domains in the dissipative quantum model of brain, Int. J. Mod. Phys. B, **14**, 853-868.
- Amari, S.I. (1985). Differential Geometrical Methods in Statistics. Springer, New York.
- Amari, S.I., Nagaoka, H. (2000). Methods of Information Geometry, Oxford Univ. Press and Amer. Math. Soc.
- Ambjørn, J., Kristjansen, C.F. (1993). Non-perturbative 2D quantum gravity and Hamiltonians unbounded from below. Int. J. Mod. Phys. A **8**, 1259-

1282.

- Ambjørn, J., Loll, R. (1998). Non-perturbative Lorentzian quantum gravity, causality and topology change. *Nucl. Phys. B* **536**, 407–434.
- Ambjørn, J., Durhuus, B., Jonsson, T. (1997). *Quantum geometry*. Cambridge Monographs on Mathematical Physics, Cambridge Univ. Press, Cambridge.
- Ambjørn, J., Carfora, M., Marzuoli, A. (1998). *The Geometry of Dynamical Triangulations*. Springer, Berlin.
- Ambjørn, J., Jurkiewicz, J., Loll, R. (2000). Lorentzian and Euclidean quantum gravity – analytical and numerical results. In *M-Theory and Quantum Geometry*, eds. L. Thorlacius and T. Jonsson, NATO Science Series, Kluwer, 382–449.
- Ambjørn, J., Jurkiewicz, J., Loll, R. (2000). A non-perturbative Lorentzian path integral for gravity. *Phys. Rev. Lett.* **85**, 924–927.
- Ambjørn, J., Jurkiewicz, J., Loll, R. (2001). Dynamically triangulating Lorentzian quantum gravity. *Nucl. Phys. B* **610**, 347–382.
- Ambjørn, J., Jurkiewicz, J., Loll, R. (2001). Nonperturbative 3D Lorentzian quantum gravity. *Phys. Rev. D* **64**, 044–011.
- Ambjørn, J., Jurkiewicz, J., Loll, R. (2001). Computer simulations of 3D Lorentzian quantum gravity. *Nucl. Phys. B* **94**, 689–692.
- Amemiya, Y. (1993). On nonlinear factor analysis. *Proc. Soc. Stat. Section. Ann. Meet. Ame. Stat. Assoc.* 290–294.
- Anderson, B.D., Arbib, M.A., Manes, E.G. (1976). *Foundations of System Theory: Finitary and Infinitary Conditions*. Lecture Notes in Economics and Mathematical Systems Theory, Springer, New York.
- Anosov, D.V., Sinai, Ya.G. (1982). Certain smooth ergodic systems. *Russ. Math. Surv.*, **22**, 103–167.
- Anosov, D. (1969). Geodesic flows on closed Riemann manifolds with negative curvature. *Proc. Steklov Inst. Math.*, **90**, Amer. Math. Soc.
- Antoni, M., Ruffo, S. (1995). Clustering and relaxation in long-range Hamiltonian dynamics. *Phys. Rev. E*, **52**, 2361–2374.
- Antoniadis, I., Gava, E., Narain, K.S., Taylor, T.R. (1994). Topological amplitudes in string theory. *Nucl. Phys. B* **413**, 162.
- Apps, R., Garwicz, M. (2005). Anatomical and physiological foundations of cerebellar information processing. *Nature Rev. Neurosci.*, **6**, 297–311.
- Arbib, M.A. (1966). Categories of (M, R) -systems. *Bull. Math. Biol.*, **28**, 511–517.
- Arbib, M. (1987). *Brains, Machines and Mathematics*. Springer, New York.
- Arbib, M. (ed.) (1998). *Handbook of Brain Theory and Neural Networks* (2nd ed.). MIT Press, Cambridge.
- Arfken, G. (1985). *Calculus of Variations*. Ch. 17 in *Mathematical Methods for Physicists* (3rd ed.) Academic Press, Orlando, FL, 925–962.
- Aringazin, A., Mikhailov, A. (1991). Matter fields in space–time with vector non-metricity. *Clas. Quant. Grav.* **8**, 1685.
- Arkin, A.P. (2001). Synthetic cell biology. *Curr. Opin. Biotech.*, **12**, 638–644.
- Arnold, V.I. (1978). *Ordinary Differential Equations*. MIT Press, Cambridge.
- Arnold, V.I. (1988). *Geometrical Methods in the Theory of Ordinary differential*

- equations. Springer, New York.
- Arnold, V.I. (1988). Geometrical Methods in the Theory of Ordinary differential equations. Springer, New York.
- Arnold, V.I. (1989). Mathematical Methods of Classical Mechanics (2nd ed). Springer, New York.
- Arnold, V.I. (1992). Catastrophe Theory. Springer, Berlin.
- Arnold, V.I. (1993). Dynamical systems. Encyclopaedia of Mathematical Sciences, Springer, Berlin.
- Ashby, N. (1997). Relativistic effects in the global positioning system. Plenary lecture on quantum gravity at the GR15 conference, Puna, India.
- Ashcraft, M.H. (1994). Human Memory and Cognition (2nd ed). HarperCollins, New York.
- Aschieri, P., Blohmann, C., Dimitrijevic, M., Meyer, F., Schupp, P., Wess, J. (2005). A gravity theory on noncommutative spaces, *Class. Quant. Grav.* **22**, 3511.
- Ashtekar, A. (1986). New variables for classical and quantum gravity, *Phys. Rev. Lett.* **57**, 2244–2247.
- Ashtekar, A. (1987). New Hamiltonian formulation of general relativity, *Phys. Rev.* **D36**, 1587–1602.
- Ashtekar, A. (1988). New perspectives in canonical gravity. Bibliopolis.
- Ashtekar, A. (1991). Lecture notes on non-perturbative canonical gravity. Advanced Series in Astrophysics and Cosmology, Vol. 6, World Scientific, Singapore.
- Ashtekar, A., Isham, C.J. (1992). Representations of the holonomy algebras of gravity and non-Abelian gauge theories. *Class. Quant. Grav.* **9**, 1433–85.
- Ashtekar, A., Lewandowski, J. (1995). Projective techniques and functional integration, *J. Math. Phys.* **36**, 2170.
- Ashtekar, A., Husain, V., Samuel, J., Rovelli, C., Smolin, L. (1989). 2+1 quantum gravity as a toy model for the 3+1 theory, *Classical and Quantum Gravity* **6**, L185.
- Ashtekar, A., Lewandowski, J., Marolf, D., Mourão, J., Thiemann, T. (1995). Quantization of diffeomorphism invariant theories of connections with local degrees of freedom, *J. Math. Phys.* **36**, 6456–6493.
- Atiyah, M.F., Hirzebruch, F. (1961). Vector bundles and homogeneous spaces. *Proc. Symp. Pure Math.* **3**, 7–38.
- Atiyah, M.F., Singer, I.M. (1963). The Index of Elliptic Operators on Compact Manifolds, *Bull. Amer. Math. Soc.* **69**, 322–433.
- Atiyah, M.F., Singer, I.M. (1968). The Index of Elliptic Operators I, II, III. *Ann. Math.* **87**, 484–604.
- Atiyah, M.F., Anderson, D.W. (1967). *K*–Theory. Benjamin, New York.
- Atiyah, M.F., Bott, R. (1984). The moment map and equivariant cohomology. *Topology*, **23**, 1–28.
- Atiyah, M.F., Segal, G.B. (1971). Exponential isomorphisms for λ –rings. *Quart. J. Math. Oxford Ser.* **22**, 371–378.
- Atiyah, M.F., Hitchin, N.J., Singer, I.M. (1978). Self Duality in four-dimensional Riemannian geometry, *Proc. R. Soc. Lond.* **A362**, 425–461;

- Atiyah, M.F. (1988). Topological quantum field theory. *Publ. Math. IHES* **68**, 175–186.
- Atiyah, M.F. (1988). New invariants of three and four dimensional manifolds. In *The Mathematical Heritage of Hermann Weyl*, eds. R. Well et al., *Proc. Symp. Pure. Math.* 48, Am. Math. Soc., Providence.
- Atiyah, M.F. (1989). *The Geometry and Physics of Knots*. Cambridge Univ. Press, Cambridge.
- Atiyah, M.F., Hitchin, N.J. (1988). *The geometry and dynamics of magnetic monopoles*. M.B. Porter Lectures. Princeton Univ. Press, Princeton, NJ.
- Atiyah, M.F., Jeffrey, L. (1990). Topological Lagrangians and Cohomology, *J. Geom. Phys.* **7**, 119.
- Atiyah, M.F. (2000). K–Theory Past and Present. [arXiv:math/0012213](https://arxiv.org/abs/math/0012213).
- Babelon, O., Viallet, C.M. (1990). Hamiltonian structures and Lax equations. *Phys. Lett. B*, **237**, 411–416.
- Baez, J.C. (1996). Spin Network States in Gauge Theory, *Adv. Math.* **117**, 53.
- Baez, J.C. (1996). Spin Networks in Nonperturbative Quantum Gravity. In Kauffman, L.H, (ed.) *The Interface of Knots and Physics*. Am. Math. Soc., Providence, Rhode Island.
- Baez, J.C. (1996). 4-Dimensional BF Theory as a Topological Quantum Field Theory, *Lett. Math. Phys.* **38**, 128.
- Baez, J. (1997). An introduction to n –categories. *7th Conference on Category Theory and Computer Science*, E. Moggi and G. Rosolini (eds), *Lecture Notes in Computer Science*, Springer, Berlin.
- Baez, J. (2001). The Meaning of Einstein’s Equation. [arXiv:gr-qc/0103044](https://arxiv.org/abs/gr-qc/0103044).
- Baez, J. (2002). Categorified gauge theory. *Lecture in the Joint Spring Meeting of the Pacific Northwest Geometry Seminar and Cascade Topology Seminar*.
- Baez, J., Dolan, J. (1995). Higher dimensional algebra and topological quantum field theory. *J. Math. Phys.* **36**, 6073–6105.
- Baez, J., Dolan, J. (1998). Higher–Dimensional Algebra III: n –categories and the Algebra of Opetopes. *Adv. Math.* **135**(2), 145–206.
- Bailey, T.N., Eastwood, M.G. (1991). Complex paraconformal manifolds: their differential geometry and twistor theory, *Forum Math.* **3**, 61–103.
- Bailey, T.N., Eastwood, M.G., Gover, A.R. (1994). The Thomas structure bundle for conformal, projective and related structures, *Rocky Mountain Journal of Math.* **24**, 1–27.
- Bakker, B.V., Smit, J. (1995). Curvature and scaling in 4D dynamical triangulation. *Nucl. Phys.* **B439**, 239.
- Bamber, D., van Santen, J.P.H. (1985). How many parameters can a model have and still be testable? *J. Math. Psych.*, **29**, 443–473.
- Banaszuk, A., Hauser, J. (1996). Approximate feedback linearization: a homotopy operator approach. *SIAM J. Cont. & Optim.*, **34**(5), 1533–1554.
- Banks, T., Fischler, W., Shenker, S.H., Susskind, L. (1997). M-theory as a matrix model: A conjecture. *Phys. Rev.* **D55**, 5112–5128.
- Barbashov B.M., Nesterenko V.V. (1990). *Introduction to the relativistic string theory*. World Scientific.
- Barbero, F. (1994). Real-polynomial formulation of general relativity in terms of

- connections, *Phys. Rev.* **D49**, 6935–6938.
- Barbero, F. (1995). Real Ashtekar Variables for Lorentzian Signature Space–Times, *Phys. Rev.* **D 51**, 5507–5510;
- Barbero, F. (1995). Reality Conditions and Ashtekar Variables: a Different Perspective, *Phys. Rev.* **D 51**, 5498–5506.
- Barbieri, A. (1997). Quantum tetrahedra and simplicial spin networks, arXiv:gr-qc/9707010.
- Barreira, L. (1996). A non-additive thermodynamic formalism and applications to dimension theory of hyperbolic dynamical systems. *Ergodic Theory Dynam. Systems* **16**, 871–927.
- Barreira, L., Pesin, Y. (2002). *Lyapunov Exponents and Smooth Ergodic Theory*, Univ. Lecture Series 23, Am. Math. Soc.
- Barret, J., Crane, L. (1997). Relativistic spin networks and quantum gravity, arXiv:gr-qc/9709028.
- Barrow-Green, J. (1997). *Poincaré and the Three Body Problem*. American Mathematical Society, Providence, RI.
- Barry, J.C. (1993). *Matrices, Monads and the Fast Fourier Transform*. Univ. Tech., Sidney.
- Basar, T., Olsder, G.J. (1995). *Dynamic Noncooperative Game Theory* (2nd ed.), Academic Press, New York.
- Bashkirov, D., Sardanashvily, G. (2004). Covariant Hamiltonian Field Theory. Path Integral Quantization. *Int. J. Theor. Phys.* **43**, 1317–1333.
- Batlle, C., Gomis, J., Gràcia, X., Pons, J.M. (1989). Noether’s Theorem and gauge transformations: application to the bosonic string and CP_2^{n-1} model. *J. Math. Phys.* **30**, 1345.
- Bauchau, O.A., Trainelli, L. (2003). The Vectorial Parameterization of Rotation. *Non. Dyn.*, **31**(1), 71–92.
- Baulieu, L., Grossman, B. (1988). Monopoles and topological field theory. *Phys. Lett. B* **214**:2, 223.
- Baum, H. (2000). Twistor and Killing spinors on Lorentzian manifolds and their relations to CR and Kaehler geometry. *Int. Congr. Diff. Geom. in memory of Alfred Gray*, Bilbao, Spain.
- Baxter, R.J. (1982). *Exactly solved models in statistical mechanics*. Academic Press.
- Belavin, A.A., Polyakov, A.M., Swartz, A.S., Tyupkin, Yu.S. (1975). SU(2) instantons discovered. *Phys. Lett. B* **59**, 85.
- Bellman, R., Kalaba, R. (1964). *Selected papers on mathematical trends in control theory*. Dover, New York.
- Bénabou, J. (1967). Introduction to bicategories. In: *Lecture Notes in Mathematics*. Springer, New York.
- Bennati, E., Rosa-Clot, M., Taddei, S. (1999). A Path Integral Approach to Derivative Security Pricing: I. Formalism and Analytical Results, *Int. Journ. Theor. Appl. Finance* **2**, 381.
- Benvenuto, N., Piazza, F. (1992). On the complex backpropagation algorithm. *IEEE Trans. Sig. Proc.*, **40**(4), 967–969.
- Berleant, D., Kuipers, B. (1992). Qualitative–Numeric Simulation with Q3, in

- Recent Advances in Qualitative Physics. eds. Boi Faltings and Peter Struss, MIT Press, Cambridge.
- Bernstein, N.A. (1935). Investigations in Biodynamics of Locomotion (in Russian). WIEM, Moscow.
- Bernstein, N.A. (1947). On the structure of motion (in Russian). Medgiz, Moscow.
- Bernstein, N.A. (1967). The Coordination and Regulation of Movement. Pergamon Press, Oxford.
- Bernstein, N. (1982). Some emergent problems of the regulation of motor acts. In: H.T.A. Whiting (Ed) Human Motor Actions: Bernstein Reassessed, North Holland, Amsterdam.
- Bershadsky, M., Cecotti, S., Ooguri, H., Vafa, C. (1994). Kodaira-Spencer theory of gravity and exact results for quantum string amplitudes. Commun. Math. Phys. **165**, 311.
- Billingsley, P. (1965). Ergodic theory and information, Wiley, New York.
- Birmingham, D., Rakowski, M., Thompson, G. (1989). BRST quantization of topological field theories. Nucl. Phys. B **315**, 577.
- Birmingham, D., Blau, M., Rakowski, M., Thompson, G. (1991). Topological field theory. Phys. Rep. **209**, 129.
- Birrel, N.D., Davies, P.C.W. (1982). Quantum fields in curved space, Cambridge Univ. Press, Cambridge, UK.
- Black, F., Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. J. Polit. Econ. **81**, 637–659.
- Blackmore, D.L., Leu, M.C. (1992). Analysis of swept volumes via Lie group and differential equations. Int. J. Rob. Res., **11**(6), 516–537.
- Blackmore, D.L., Prykarpatsky, Y.A., Samulyak, R.V. (1998). The Integrability of Lie-invariant Geometric Objects Generated by Ideals in the Grassmann Algebra. J. Nonlin. Math. Phys., **5**(1), 54–67.
- Blanchard, A. (1956). Sur les variétés analytiques complexes. Ann. Sci. Ec. Norm. Sup. **3**, 73.
- Blattner, R. (1983). Nonlinear Partial Differential Operators and Quantization Procedure, in Proceedings, Clausthall 1981, Springer-Verlag, New York, 209–241.
- Bogoyavlensky, O. (1981). Toppling Solitons. Nonlinear Integrable Equations, Moscow, Nauka.
- Bontempi, G. (1995). Modelling with uncertainty in continuous dynamical systems: the probability and possibility approach. IRIDIA - ULB Technical Report, 95–16.
- Boos, B., Bleecker, D.D. (1985). Topology and Analysis: The Atiyah–Singer Index Formula and Gauge–Theoretic Physics, Springer, Berlin.
- Boothby, W.M. (1986). An Introduction to Differentiable Manifolds and Riemannian Geometry, Academic Press, New York.
- Bott, R. (1959). The Stable Homotopy of the Classical Groups, Ann. Math. **70**, 313–337.
- Bott, R., Tu, L.W. (1982). Differential Forms in Algebraic Topology. Graduate Texts in Mathematics, Springer, New York.
- Bouwknegt, P., Mathai, V. (2000). D-Branes, B-Fields, and Twisted K-Theory.

JHEP **0003:007**.

- Bransford, J.D., Franks, J.J. (1971). The Abstraction of Linguistic Ideas. *Cogn. Psych.*, **2**, 331–350.
- Breckenridge, J.C., Myers, R.C., Peet, A.W., Vafa, C. (1997). D-branes and spinning black holes. *Phys. Lett. B* **391**, 93.
- Breitenbach, G., Schiller, S., Mlynek, J. (1997). Measurement of the quantum states of squeezed light. *Nature*, **387**, 471–475.
- Broadbent, D.E. (1958). *Perception and communications*. Pergamon Press, London.
- Brockett, R. (2001). New Issues in the Mathematics of Control. In *Mathematics Unlimited – 2001 and Beyond*, Springer, New York.
- van den Broeck, C., Parrondo, J.M.R., Toral, R. (1994). Noise-Induced Nonequilibrium Phase Transition. *Phys. Rev. Lett.* **73**, 3395–3398.
- van den Broeck, C., Parrondo, J.M.R., Toral, R., Kawai, R. (1997). Nonequilibrium phase transitions induced by multiplicative noise. *Phys. Rev. E* **55**, 4084–4094.
- Bryant, R., Chern, S., Gardner, R., Goldschmidt, H., Griffiths, P. (1991). *Exterior Differential Systems*. Springer, Berlin.
- Bryant, R., Griffiths, P., Grossman, D. (2003). *Exterior Differential Systems and Euler–Lagrange Partial differential Equations*. Univ. Chicago Press, Chicago.
- Busemeyer, J.R., Diederich, A., (2002). Survey of decision field theory. *Math. Soc. Sci.*, **43**, 345–370.
- Busemeyer, J.R., Townsend, J.T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psych. Rev.*, **100**, 432–459.
- Chari, V., Pressley, A. (1995). *A Guide to Quantum Groups*, Cambridge Univ. Press, Cambridge, UK.
- Caiani, L., Casetti, L., Clementi, C., Pettini, M. (1997). Geometry of dynamics, Lyapunov exponents and phase transitions. *Phys. Rev. Lett.* **79**, 4361.
- Calabi, E. (1957). On Kähler manifolds with vanishing canonical class. In *Algebraic geometry and topology: a symposium in honor of S. Lefschetz*, eds. R.H. Fox, D.C. Spencer and A.W. Tucker, Princeton Univ. Press.
- Calogero, F. (1975). Exactly solvable one-dimensional many-body problems, *Lett. Nuovo Cim.* **13**, 411–416.
- Canarutto, D. (1986). Bundle splittings, connections and locally principle fibred manifolds. *Bull. U.M.I. Algebra e Geometria, Seria VI* **V-D**, 18.
- Candelas, P., Horowitz, G.T., Strominger, A., Witten, E. (1985). Vacuum configurations for superstrings, *Nucl. Phys. B* **258**, 46.
- Cao, H.D., Chow, B. (1999). Recent Developments on the Ricci Flow. *Bull. Amer. Math. Soc.* **36**, 59–74.
- Cariñena, J., Fernández-Núñez, J. (1993). Geometric theory of time-dependent singular Lagrangians, *Fortschr. Phys.* **41**, 517.
- Cariñena, J., Rañada, M. (1989). Poisson maps and canonical transformations for time-dependent Hamiltonian systems. *J. Math. Phys.* **30**, 2258.
- Cariñena, J., Rañada, M. (1993). Lagrangian systems with constraints. A geo-

- metric approach to the method of Lagrange multipliers. *J. Phys. A*. **26**, 1335.
- Cariñena, J., Crampin, M., Ibort, L. (1991). On the multisymplectic formalism for first order field theories. *Diff. Geom. Appl.* **1**, 345.
- Cariñena, J., Gomis, J., Ibort, L. and Román, N. (1995). Canonical transformation theory for presymplectic systems. *J. Math. Phys.* **26**, 1961.
- Carlip, S., Nelson, J.E. (1995). Comparative Quantizations of (2+1)-Dimensional Gravity. *Phys. Rev.* **D51**, 5643.
- Carpenter, G.A., Grossberg, S. (2003). Adaptive Resonance Theory. In M.A. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks*, Second Edition, Cambridge, MA: MIT Press, 87–90.
- Casetti, L., Cohen, E.G.D., Pettini, M. (1999). Origin of the Phase Transition in a Mean-Field Model. *Phys. Rev. Lett.*, **82**, 4160.
- Casetti, L., Pettini, M., Cohen, E.G.D. (2003). Phase transitions and topology changes in configuration space. *J. Stat. Phys.*, **111**, 1091.
- Cecotti, S., Vafa, C. (1991). Topological – antitopological fusion, *Nucl. Phys. B* **367**, 359.
- Celeghini, E., Rasetti, M., Vitiello, G. (1992). Quantum Dissipation. *Annals Phys.*, **215**, 156.
- Chaichian, M., Kulish, P., Nishijima, K., Tureanu, A. (2004). On a Lorentz-invariant interpretation of noncommutative space-time and its implications on noncommutative QFT, *Phys. Lett.* **B604**, 98.
- Chamseddine, A.H., Connes, A. (1996). The Spectral Action Principle. *Phys. Rev. Lett.* **24**, 4868-4871.
- Chen, G., Dong, X. (1998). *From Chaos to Order. Methodologies, Perspectives and Application*. World Scientific, Singapore.
- Cheng, T.-P., Li, L.-F. (1984). *Gauge Theory of Elementary Particle Physics*. Clarendon Press, Oxford.
- Chern, S.S., (1946). Characteristic classes of Hermitian manifolds, *Ann. of Math.* **47**, 85-121.
- Chern, S.S. (1996). Riemannian Geometry as a Special Case of Finsler Geometry, *Cont. Math.*, 51–58, Vol. 196, Amer. Math. Soc. Providence, RI.
- Chevalley, C. (1955). *Theorie differential equations groupes de Lie*. vol. 1–3. Hermann D.C., Paris.
- Chinea, D., de León, M., Marrero, J. (1994). The constraint algorithm for time-dependent Lagrangians, *J. Math. Phys.* **35**, 3410.
- Cho, M.W., Kim, S. (2004). Understanding visual map formation through vortex dynamics of spin Hamiltonian models. *Phys. Rev. Lett.* **92**(1), 18101.
- Choquet-Bruhat, Y., DeWitt-Morete, C. (1982). *Analysis, Manifolds and Physics* (2nd ed). North-Holland, Amsterdam.
- Choquet-Bruhat, Y., DeWitt-Morete, C. (2000). *Analysis, Manifolds and Physics, Part II: 92 Applications* (rev. ed). North-Holland, Amsterdam.
- Choptuik, M.W. (1993). Universality and scaling in gravitational collapse of a massless scalar field. *Phys. Rev. Lett.*, **70**, 9–12.
- Chu, C.-S., Ho, P.-M. (1999). Noncommutative Open String and D-brane, *Nucl. Phys.* **B550**, 151.

- Clementi, C. Pettini, M. (2002). A geometric interpretation of integrable motions. *Celest. Mech. & Dyn. Astr.*, **84**, 263–281.
- Cohen, M.A., Grossberg, S. (1983). Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Trans. Syst., Man, Cybern.*, **13**(5), 815–826.
- Coleman, S. (1988). *Aspects of Symmetry: Selected Erice Lectures*. Cambridge Univ. Press, Cambridge.
- Collins, A.M., Quillian, M.R. (1969). Retrieval Time From Semantic Memory, *J. Verb. Learn. & Verb. Behav.*, **8**, 240–248.
- Connes, A. (1994). *Noncommutative Geometry*, Academic Press, New York.
- Connes, A., Lott, J. (1990). Particle models and non-commutative geometry. *Nucl. Phys.* **B18**, 29–47.
- Connes, A., Douglas, M.R., Schwarz, A. (1997). Noncommutative Geometry and Matrix Theory: Compactification on Tori, arXiv:hep-th/9711162.
- Contreras-Vidal, J.L., Grossberg, S., Bullock, D. (1997). A neural model of cerebellar learning for arm movement control: Cortico-spino-cerebellar dynamics. *Learning and Memory*, **3**, 475–502.
- Conway, E.D., Hopf, E. (1964). Hamilton's Theory and Generalized Solutions of the Hamilton–Jacobi Equation. *J. Math. Mech.* **13**, 939–986.
- Conway, J.H., Curtis, R.T., Norton, S.P., Parker, R.A., Wilson, R.A. (1985). *Atlas of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups*. Clarendon Press, Oxford.
- Cox, E. (1992). Fuzzy Fundamentals, *IEEE Spectrum*, 58–61.
- Cox, E. (1994). *The Fuzzy Systems Handbook*. AP Professional.
- Craik, F., Lockhart, R. (1972). Levels of processing: A framework for memory research. *J. Verb. Learn. & Verb. Behav.*, **11**, 671–684.
- Crane, L., Yetter, D. (1993). A categorical construction of 4D topological quantum field theories. In *Quantum Topology*, R Baadhio and LH Kauffman ed. World Scientific, Singapore.
- Crane, L., Frenkel, I. (1994). Four dimensional topological quantum field theory, Hopf categories, and the canonical bases. *Jour. Math. Phys.* **35**, 5136–5154.
- Crawford, J. (1991). Clifford algebra: Notes on the spinor metric and Lorentz, Poincaré and conformal groups. *J. Math. Phys.* **32**, 576.
- Cremmer, E., Schwimmer, A., Thorn, C. (1986). The Vertex Function in Witten's Formulation of String Field Theory, *Phys. Lett.* **B179**, 57.
- Crutchfield, J.P., Young, K. (1989). Computation at the onset of chaos. In *Complexity, Entropy and the Physics of Information*, pag. 223, SFI Studies in the Sciences Complexity, vol. VIII, W.H. Zurek (Ed.), Addison-Wesley, Reading.
- D'Hoker, E., Krichever, I.M., Phong, D.H. (1997). The effective prepotential for $N = 2$ supersymmetric $SU(N_c)$ gauge theories, *Nucl. Phys.* **B489**, 179.
- D'Hoker, E., Phong, D.H., (1998). Calogero-Moser systems in $SU(N)$ Seiberg-Witten theory, *Nucl. Phys.* **B513**, 405–444.
- D'Hoker, E., Phong, D.H., (1998). Calogero-Moser Lax pairs with spectral parameter for general Lie algebras, *Nucl. Phys.* **B530**, 537–610.
- D'Hoker, E., Phong, D.H., (1998). Calogero-Moser and Toda systems for twisted

- and untwisted affine Lie algebras, Nucl. Phys. **B530**, 611-640.
- D'Hoker, E., Phong, D.H., (1998). Spectral curves for super Yang-Mills with adjoint hypermultiplet for general Lie algebras, Nucl. Phys. **B534**, 697-719
- Dasgupta, A. (2002). The real Wick rotations in quantum gravity. JHEP, **0207**, 062.
- Dasgupta, A., Loll, R. (2001). A proper-time cure for the conformal sickness in quantum gravity. Nucl. Phys. B **606**, 357-379.
- Dauxois, T., Theodorakopoulos, N., Peyrard, M. (2002). Thermodynamic instabilities in one dimension: correlations, scaling and solitons. J. Stat. Phys. **107**, 869.
- Davies, E.B. (1989). Heat Kernels and Spectral Theory. Cambridge Univ. Press.
- Davydov, A.S. (1981). Biology and Quantum Mechanics, Pergamon Press, New York.
- Davydov, A.S. (1991). Solitons in Molecular Systems. (2nd ed), Kluwer, Dordrecht.
- Dearnaley, R. (1990). The Zero-Slope Limit of Witten's String Field Theory with Chan-Paton Factors, Nucl. Phys. **B334**, 217.
- Deligne, P., Etingof, P., Freed, D.S., Jeffrey, L.C., Kazhdan, D., Morgan, J.W., Morrison, D.R., Witten, E. (1999). Quantum Fields and Strings: A Course for Mathematicians, Am. Math. Soc.
- DePietri, R. (1997). On the relation between the connection and the loop representation of quantum gravity, Class. and Quantum Grav. **14**, 53-69.
- DePietri, R., Rovelli, C. (1996). Geometry Eigenvalues and Scalar Product from Recoupling Theory in Loop Quantum Gravity, Phys. Rev. **D54**, 2664-2690.
- Dieudonne, J.A. (1969). Foundations of Modern Analysis (in four volumes). Academic Press, New York.
- Dieudonne, J.A. (1988). A History of Algebraic and Differential Topology 1900-1960. Birkhäuser, Basel.
- Dighton, K. (1974). An introduction to the theory of local twistors, Int. J. Theor. Phys. **11**, 31-43.
- Dijkgraaf, R., Verlinde, H., Verlinde, E. (1991). Notes On Topological String Theory And 2D Quantum Gravity, in String Theory and Quantum Gravity, Proceedings of the Trieste Spring School 1990, (eds.) M. Green et al., World Scientific, 91-156.
- Dimitrijevic, Möller, L., Tsouchnika, E., Wess, J., Wohlgenannt, M. (2003). Deformed field theory on kappa-spacetime. Eur. Phys. J. **C31**, 129.
- Dimitrijevic, M., Meyer, F., Möller, L., Wess, J. (2004). Gauge theories on the kappa-Minkowski spacetime. Eur. Phys. J. **C36**, 117.
- Dirac, P.A.M. (1982). Principles of Quantum Mechanics (4th ed.). Oxford Univ. Press, Oxford.
- Dittrich, W., Reuter, M. (1994). Classical and Quantum Dynamics. Springer Verlag, Berlin.
- Dodson, C.T.J., Parker, P.E. (1997). A User's Guide to Algebraic Topology. Kluwer, Dordrecht.
- Dolgopyat, D., Pesin, Ya. (2002). Every compact manifold carries a completely hyperbolic diffeomorphism. Ergodic Theory Dynam. Systems **22**, 409-435.

- Domany, E., Van Hemmen, J.L., Schulten, K. (eds.) (1991). *Models of Neural Networks*. Springer, Berlin.
- De Dominicis, C. (1978). Dynamics as a substitute for replicas in systems with quenched random impurities. *Phys. Rev. B* **18**, 4913–4919.
- Donagi, R., Witten, E. (1996). Supersymmetric Yang–Mills and integrable systems, *Nucl. Phys.* **B460**, 288–334.
- Donaldson, S. (1986). Connections, cohomology, and the intersection forms of four-manifolds, *J. Diff. Geom.* **24**, 275–341;
- Donaldson, S. (1987). The Orientation of Yang–Mills Moduli Spaces and Four-Manifold Topology. *J. Diff. Geom.* **26**, 397.
- Donaldson, S. (1990). Polynomial invariants for smooth four-manifolds, *Topology* **29**, 257–341;
- Donaldson, S., Kronheimer, P. (1990). *The Geometry of Four-Manifolds*, Oxford.
- Donoghue, J.F. (1996). The Quantum Theory of General Relativity at Low Energies. *Helv. Phys. Acta.* **69**, 269–275.
- Doplicher, S., Fredenhagen, K., Roberts, J.E. (1994). Space-time quantization induced by classical gravity. *Phys. Lett.* **B331**, 39–44.
- Dorogovtsev, S.N., Mendes, J.F.F. (2003). *Evolution of Networks*. Oxford Univ. Press.
- Douglas, M.R., Nekrasov, N.A. (2001). Noncommutative field theory. *Rev. Mod. Phys.* **73**, 977.
- Dowker, F. (2002). Topology change in quantum gravity. In *Proceedings of Stephen Hawking’s 60th birthday conference*, Cambridge Univ Press.
- Dragovic, V., Gajic, B. (2003). The Wagner Curvature Tensor in Nonholonomic Mechanics. *Reg. Chaot. Dyn.*, **8**(1), 105–124.
- Drinfeld, V.G. (1985). Quantum groups, *Proc. ICM*, Berkeley, 798.
- Dubois, D., Prade, H. (1980). *Fuzzy Sets and Systems*. Academic Press, New York.
- Dubrovin, B., Krichever, I., Novikov, S. (1985). Integrable systems – I. *Sovremennye problemy matematiki (VINITI)*, Dynamical systems **4**, 179.
- Dunne, G.V. (1999). Aspects of Chern–Simons Theory. arXiv: hep-th/9902115.
- Dustin, P. (1984). *Microtubules*. Springer, Berlin.
- Eastwood, M.G., Graham, C.R. (1991). Invariants of conformal densities. *Duke Math. Jour.* **63**, 633–671.
- Eastwood, M. (2002). Higher symmetries of the Laplacian. arXiv:hep-th/0206233.
- Eccles, J.C. (1964). *The Physiology of Synapses*. Springer, Berlin.
- Eccles, J.C., Ito, M., Szentagothai, J. (1967). *The Cerebellum as a Neuronal Machine*. Springer, Berlin.
- Echeverría-Enríquez, A., Muñoz-Lecanda, M., Román-Roy, N. (1991). Geometrical setting of time-dependent regular systems. *Alternative models*, *Rev. Math. Phys.* **3**, 301.
- Echeverría-Enríquez, A., Muñoz-Lecanda, M., Román-Roy, N. (1995). Non-standard connections in classical mechanics. *J. Phys. A* **28**, 5553.
- Echeverría-Enríquez, A., Muñoz-Lecanda, M., Román-Roy, N. (1998). Multivector-fields and connections. Setting Lagrangian equations in field theories. *J. Math. Phys.* **39**, 4578.

- Eells, J., Lemaire, L. (1978). A report on harmonic maps. *Bull. London Math. Soc.* **10**, 1-68.
- Elkin, V.I. (1999). *Reduction of Nonlinear Control Systems. A Differential Geometric Approach*. Kluwer, Dordrecht.
- Ellis, J., Mavromatos, N., Nanopoulos, D.V. (1992). String theory modifies quantum mechanics. *CERN-TH/6595*.
- Ellis, J., Mavromatos, N. and Nanopoulos, D.V. (1999). A microscopic Liouville arrow of time. *Chaos, Solit. Fract.*, **10**(2-3), 345-363.
- Elworthy, K.D. (1982). *Stochastic Differential Equations on Manifolds*. Cambridge Univ. Press, Cambridge.
- Fabinger, M., Horava, P. (2000). Casimir Effect Between World-Branes in Heterotic M -Theory. *Nucl. Phys.* **B580**, 243.
- Fatibene, L., Ferraris, M., Francaviglia, M. (1994). Nöther formalism for conserved quantities in classical gauge field theories II. *J. Math. Phys.*, **35**, 1644.
- Federer, H. (1969). *Geometric Measure Theory*. Springer, New York.
- Ferraris, M., Kijowski, J. (1982). On the equivalence of the relativistic theories of gravitation. *GRG* **14**, 165.
- Feynman, R.P. (1948). Space-time approach to non-relativistic quantum mechanics. *Rev. Mod. Phys.* **20**, 367-387.
- Feynman, R.P. (1951). *An Operator Calculus Having Applications in Quantum Electrodynamics*. *Phys. Rev.* **84**, 108-128.
- Feynman, R.P. (1972). *Statistical Mechanics, A Set of Lectures*. WA Benjamin, Inc., Reading, Massachusetts.
- Feynman, R.P. (1998). *Quantum Electrodynamics*. Advanced Book Classics, Perseus Publishing.
- Feynman, R.P., Hibbs, A.R. (1965). *Quantum Mechanics and Path Integrals*, McGraw-Hill, New York.
- Fine, A. (1986). *The Shaky Game: Einstein, Realism and Quantum Theory*, Univ. Chicago Press, Chicago.
- Flanders, H. (1963). *Differential Forms with Applications to the Physical Sciences*. Acad. Press, New York.
- Floer, A. (1987). Morse theory for fixed points of symplectic diffeomorphisms. *Bull. AMS.* **16**, 279.
- Floer, A. (1988). Morse theory for Lagrangian intersections. *J. Diff. Geom.*, **28**(9), 513-517.
- Flume, R., O'Raifeartaigh, L., Sachs, I. (1996). Brief Resume of Seiberg-Witten Theory. *DIAS-STP* 96-22.
- Fordy, A.P. (ed.) (1990). *Soliton Theory: A Survey of Results*. MUP, Manchester.
- Forger, M., Römer, H. (2004). Currents and the energy-momentum tensor in classical field theory: a fresh look at an old problem. *Ann. Phys. (N.Y.)* **309**, 306-389.
- Forger, M., Paufler, C., Römer, H. (2004). *Hamiltonian Multivector Fields and Poisson Forms in Multisymplectic Field Theory*. arXiv:math-ph/0407057.
- Fox, C. (1988). *An Introduction to the Calculus of Variations*. Dover, New York.
- Franzosi, R., Casetti, L., Spinelli, L., Pettini, M. (1999). Topological aspects of

- geometrical signatures of phase transitions. *Phys. Rev. E*, **60**, 5009–5012.
- Franzosi, R., Pettini, M. Spinelli, L. (2000). Topology and phase transitions: a paradigmatic evidence. *Phys. Rev. Lett.* 84(13), 2774–2777.
- Franzosi, R., Pettini, M. (2004). Theorem on the origin of phase transitions. *Phys. Rev. Lett.*, **92**(6), 60601.
- Freed, D., Uhlenbeck, K.K. (1984). *Instantons and four manifolds*, Springer, New York.
- Freed, D.S., (2001). The Verlinde Algebra is Twisted Equivariant K–Theory. *Turk. J. Math* **25**, 159–167.
- Friberg, O. (1988). A Set of Parameters for Finite Rotations and Translations. *Comp. Met. App. Mech. Eng.*, **66**, 163–171.
- Frittelli, S., Kozameh, C., Newman, E.T. (1995). General Relativity via Characteristic Surfaces. *J. Math. Phys.* **36**, 4984.
- Freed, D., Hopkins, M. Teleman, C. (2003). Twisted K–theory and loop group representations, arXiv:math.AT/0312155.
- Frittelli, S., Kozameh, C., Newman, E.T., Rovelli, C., Tate, R.S. (1997). Fuzzy space–time points from the null–surface formulation of general relativity, *Class. Quantum Gravity*, **14**, A143.
- Frölich, H., Kremer, F. (1983). *Coherent Excitations in Biological Systems*. Springer, New York.
- Fröhlich, J., Gawedzki, K. (1994). Conformal field theory and the geometry of strings. *CRM Proceedings and Lecture Notes* **7**, 57–97.
- Fulling, S. (1989). *Aspects of quantum field theory in curved space–time*, Cambridge Univ. Press.
- Fulton, W., Harris, J. (1991). *Representation Theory. A First Course*, GTM 129, RIM, Springer, New York.
- Gabbiani, F., Midtgaard, J., and Knoepfl, T. (1994). Synaptic integration in a model of cerebellar granule cells. *J. Neurophysiol.*, **72**, 999–1009.
- Gómez, J.C. (1994). Using symbolic computation for the computer aided design of nonlinear (adaptive) control systems. Tech. Rep. EE9454, Dept. Electr. and Comput. Eng., Univ. Newcastle, Callaghan, NSW, AUS.
- Gallavotti, G. (1983). *The Elements of Mechanics*. Springer-Verlag, Berlin.
- García, P. (1972). Connections and 1-jet fibre bundles. *Rend. Sem. Univ. Padova*, **47**, 227.
- García, P. (1977). Gauge algebras, curvature and symplectic structure. *J. Diff. Geom.*, **12**, 209.
- Gardiner, C.W. (1985). *Handbook of Stochastic Methods for Physics, Chemistry and Natural Sciences* (2nd ed). Springer, Berlin.
- Georgii, H.O. (1988). *Gibbs Measures and Phase Transitions*. Walter de Gruyter, Berlin.
- Georgiou, G.M., Koutsougeras, C. (1992). Complex domain backpropagation, *IEEE Trans. Circ. Sys.*, **39**(5), 330–334.
- Ghez, C. (1990). Introduction to motor system. In: Kandel, E.K. and Schwarz, J.H. (eds.) *Principles of neural science*. 2nd ed. Elsevier, Amsterdam, 429–442.
- Ghez, C. (1991). Muscles: Effectors of the Motor Systems. In: *Principles of Neural*

- Science. 3rd Ed. (Eds. E.R. Kandel, J.H. Schwartz, T.M. Jessell), Appleton and Lange, Elsevier, 548–563.
- Giachetta, G. (1992). Jet manifolds in nonholonomic mechanics. *J. Math. Phys.* **33**, 1652.
- Giachetta, G., Mangiarotti, L. (1990). Gauge-invariant and covariant operators in gauge theories. *Int. J. Theor. Phys.*, **29**, 789.
- Giachetta, G., Sardanashvily, G. (1995). Stress-energy–momentum of affine–metric gravity. Generalized Komar superpotential. arXiv:gr-qc/9511008.
- Giachetta, G., Sardanashvily, G. (1996). Stress-Energy-Momentum of Affine-Metric Gravity. *Class.Quant.Grav.* **13** L, 67–72.
- Giachetta, G., Sardanashvily, G. (1997). Dirac Equation in Gauge and Affine-Metric Gravitation Theories. *Int. J. Theor. Phys.* **36**, 125–142.
- Giachetta, G., Mangiarotti, L., Sardanashvily, G. (1997). *New Lagrangian and Hamiltonian Methods in Field Theory*, World Scientific, Singapore.
- Giachetta, G., Mangiarotti, L., Sardanashvily, G. (1999). Covariant Hamilton equations for field theory. *J. Phys. A* **32**, 6629.
- Giachetta, G., Mangiarotti, L., Sardanashvily, G. (2002). Covariant geometric quantization of non-relativistic Hamiltonian mechanics. *J. Math. Phys.* **43**, 56.
- Giachetta, G., Mangiarotti, L., Sardanashvily, G. (2002). Geometric quantization of mechanical systems with time-dependent parameters. *J. Math. Phys.*, **43**, 2882.
- Giachetta, G., Mangiarotti, L., Sardanashvily, G. (2005). Lagrangian supersymmetries depending on derivatives. *Global analysis and cohomology. Commun. Math. Phys.* **259**, 103–128.
- Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall.
- Glauber, R.J. (1963). The Quantum Theory of Optical Coherence. *Phys. Rev.* **130**, 2529–2539.
- Glauber, R.J. (1963). Coherent and Incoherent States of the Radiation Field. *Phys. Rev.* **131**, 2766–2788.
- Gold, M. (1999). *A Kurt Lewin Reader, the Complete Social Scientist*. Am. Psych. Assoc., Washington.
- Golovnev, A.V., Prokhorov, L.V. (2005). On Dynamics of Strings and Branes. arXiv:hep-th/0504227.
- Goodwine, B., Burdick, J. (1997). Trajectory generation for kinematic legged robots. In *IEEE Int. Conf. Rob. Aut.*, 2689–2696. IEEE Press.
- Goodwine, J.W. (1998). *Control of Stratified Systems with Robotic Applications*. PhD thesis, Caltech, Pasadena.
- Gordon, R., Power, A.J., Street, R. (1995). Coherence for tricategories. *Memoirs Amer. Math. Soc.* **117**(558).
- Gorskii, A., Krichever, I.M., Marshakov, A., Mironov, A., Morozov, A. (1995). Integrability and Seiberg-Witten exact solution, *Phys. Lett.* **B355**, 466.
- Gotay, M., Lashof, R., Sniatycki, J., Weinstein, A. (1983). Closed forms on symplectic fiber bundles, *Comm. Math. Helv.* **58**, 617–621.
- Gotay, M., Nester, J., Hinds, G. (1978). Presymplectic manifolds and the Dirac-

- Bergmann theory of constraints. *J. Math. Phys.* **19**, 2388.
- Gotay, M.J., Nester, J.M. (1980). Generalized constraint algorithm and special presymplectic manifolds. In G. E. Kaiser, J. E. Marsden (eds.), *Geometric methods in mathematical physics*, Proc. NSF-CBMS Conf., Lowell/Mass. 1979, Berlin: Springer-Verlag, Lect. Notes Math. **775**, 78–80.
- Gotay, M. (1982). On coisotropic imbeddings of presymplectic manifolds. *Proc. Amer. Math. Soc.* **84**, 111.
- Gotay, M.J. (1991). A multisymplectic framework for classical field theory and the calculus of variations. I: Covariant Hamiltonian formalism. In M. Francaviglia (ed.), *Mechanics, analysis and geometry: 200 years after Lagrangian*. North-Holland, Amsterdam, 203–235.
- Gotay, M.J. (1991). A multisymplectic framework for classical field theory and the calculus of variations. II: Space + time decomposition. *Differ. Geom. Appl.* **1**(4), 375–390.
- Gotay, M.J., Marsden, J.E. (1992). Stress-energy-momentum tensors and the Belinfante-Rosenfeld formula. *Contemp. Math.* **132**, AMS, Providence, 367–392.
- Gotay, M.J., Isenberg, J., Marsden, J.E. (1998). Momentum Maps and the Hamiltonian Structure of Classical Relativistic Fields. arXiv:hep/9801019.
- Gotay, M.J., Isenberg, J., Marsden, J.E., Montgomery, R. (2004). Momentum Maps and Classical Relativistic Fields II: Canonical Analysis of Field Theories. arXiv:math-ph/0411032.
- Gottlieb, D. (1975). Witnesses, transgressions and the evaluation map, *Indiana Univ Math. J.* **24**, 825–836.
- Goto, T. (1971). Relativistic quantum mechanics of one-dimensional mechanical continuum and subsidiary condition of dual resonance model. *Prog. Theor. Phys.* **46**, 1560.
- Gover, A.R. Slovák, J. (1999). Invariant local twistor calculus for quaternionic structures and related geometries, *J. Geom. Physics*, **32**(1), 14–56.
- Graham, R. (1978). Path-integral methods on nonequilibrium thermodynamics and statistics. In *Stochastic Processes in Nonequilibrium Systems*, (Edited by L. Garrido, P. Seglar, and P.J. Shepherd), pp. 82–138, Springer, New York.
- Grebogi, C., Ott, E., Yorke, J.A. (1987). Chaos, strange attractors, and fractal basin boundaries in nonlinear dynamics. *Science*, **238**, 632–637.
- Green, M.B., Schwarz, J.H., Witten, E. (1987). *Superstring Theory*, 2 Vols., Cambridge Univ. Press, Cambridge.
- Greene, B.R. (1996). String Theory on Calabi-Yau Manifolds. Lectures given at the TASI-96 summer school on Strings, Fields and Duality.
- Greene, B.R. (2000). *The Elegant Universe: Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory*. Random House.
- Griffiths, P.A. (1983). *Exterior Differential Systems and the Calculus of Variations*, Birkhauser, Boston.
- Griffiths, P.A. (1983). Infinitesimal variations of Hodge structure. III. Determinantal varieties and the infinitesimal invariant of normal functions. *Compositio Math.*, **50**(2-3), 267–324.

- Grinza, P., Mossa, A. (2004). Topological origin of the phase transition in a model of DNA denaturation. *Phys. Rev. Lett.* **92**(15), 158102.
- van Groesen, E., de Jager, E.M. (eds.) (1994). *Mathematical Structures in Continuous Dynamical Systems*. Studies in Mathematical Physics, vol. 6 North-Holland, Amsterdam.
- Gromov, M., Lawson, B. (1980). The classification of simply connected manifolds of positive scalar curvature, *Ann. Math.* **111**, 423.
- Grosche, C., Steiner, F. (1998). *Handbook of Feynman path integrals*. Springer Tracts in Modern Physics 145, Springer, Berlin.
- Grossberg, M., Karshon, Y. (1994). Bott towers, complete integrability and the extended character of representations. *Duke Math. Journ.* **76**, 23–58.
- D.J. Gross and A. Jevicki (1987). Operator Formulation of Interacting String Field Theory (I), *Nucl. Phys.* **B283**, 1.
- D.J. Gross and A. Jevicki (1987). Operator Formulation of Interacting String Field Theory (II), *Nucl. Phys.* **B287**, 225.
- Grossberg, S. (1969). Embedding fields: A theory of learning with physiological implications. *J. Math. Psych.* **6**, 209–239.
- Grossberg, S. (1982). *Studies of Mind and Brain*. Dordrecht, Holland.
- Grossberg, S. (1988). *Neural Networks and Natural Intelligence*. MIT Press, Cambridge.
- Grossberg, S. (1999). How does the cerebral cortex work? Learning, attention and grouping by the laminar circuits of visual cortex. *Spatial Vision* **12**, 163–186.
- Grothendieck, A. (1983). Pursuing stacks (Unpublished manuscript, distributed from UCNW). Bangor, UK.
- Grunwald, P. (1999). Viewing all models as ‘probabilistic’. Proceedings of the Twelfth Annual Conference on Computational Learning Theory (COLT’99), Santa Cruz, CA.
- Grunwald, P. (2000). The minimum description length principle. *J. Math. Psych.*, **44**, 133–152.
- Grunwald, P., Myung, I.J., Pitt, M.A. (eds.) (2005). *Advances in Minimum Description Length: Theory and Applications*. MIT Press, Cambridge, MA.
- Guet, C., Elowitz, M.B., Hsing, W., Leibler, S. (2002). Combinatorial synthesis of genetic networks. *Science*, **296**(5572), 1466–1470.
- Guillemin, V., Lerman, E., Sternberg, S. (1998). *Symplectic fibrations and multiplicity diagrams*. Cambridge Univ. Press, Cambridge.
- Gunion, J.F. (2003). *Class Notes on Path-Integral Methods*. U.C. Davis, 230B.
- Gutzwiller, M. (1990). *Chaos in Classical and Quantum Mechanics*. Springer, Berlin.
- Haag, R. (1992). *Local Quantum Physics*. Springer, Berlin.
- Haken, H. (1983). *Synergetics: An Introduction* (3rd ed). Springer, Berlin.
- Haken, H. (1993). *Advanced Synergetics: Instability Hierarchies of Self-Organizing Systems and Devices* (3rd ed.). Springer, Berlin.
- Haken, H. (1996). *Principles of Brain Functioning: A Synergetic Approach to Brain Activity, Behavior and Cognition*, Springer, Berlin.
- Haken, H. (2000). *Information and Self-Organization: A Macroscopic Approach*

- to Complex Systems. Springer, Berlin.
- Haken, H. (2002). *Brain Dynamics, Synchronization and Activity Patterns in Pulse-Codupled Neural Nets with Delays and Noise*, Springer, New York.
- Hale, J.K., Lunel, S.M.V. (1993). *Introduction to Functional Differential Equations*. Springer, New York.
- Hameroff, S.R. (1987). *Ultimate Computing: Biomolecular Consciousness and Nanotechnology*. North-Holland, Amsterdam.
- Hameroff, S.R., Penrose, R. (1996). Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness. In: Hameroff, S. R., Kaszniak, A.W. and Scott, A.C. Eds: *Toward a Science of Consciousness: the First Tucson Discussion and Debates*, 507–539. MIT Press, Cambridge.
- Hamilton, R.S. (1982). Three-manifolds with positive Ricci curvature, *J. Differential Geom.* **17**, 255–306.
- Hamoui, A., Lichnerowicz, A. (1984). Geometry of dynamical systems with time-dependent constraints and time-dependent Hamiltonians: An approach towards quantization. *J. Math. Phys.* **25**, 923.
- Hartle, J.B., Williams, R.M., Miller, W.A., Williams, R. (1997). Signature of the Simplicial Supermetric, *Class. Quant. Grav.* **14**, 2137–2155.
- Hat, N.G. (1996). Coupling the neural and physical dynamics in rhythmic movements. *Neural Computation*, **8**, 567–581.
- Hatcher, A. (2002). *Algebraic Topology*. Cambridge Univ. Press, Cambridge.
- Hatze, H. (1977). A myocybernetic control model of skeletal muscle. *Biol. Cyber.* **25**, 103–119.
- Hatze, H. (1977). A complete set of control equations for the human musculoskeletal system. *J. Biomech.* **10**, 799–805.
- Hatze, H. (1978). A general myocybernetic control model of skeletal muscle. *Biol. Cyber.*, **28**, 143–157.
- Hatze, H. (1980). Neuromusculoskeletal Control Systems Modelling — a Critical Survey of Recent Developments. *IEEE Trans. Aut. Con.* **25**(3), 375–385.
- Hatze, H. (1981). A comprehensive model for human motion simulation and its application to the take-off phase of the long jump. *J. Biomech.* **14**, 135–142.
- Hauser, J., Sastry, S., Kokotovic, P. (1992). Nonlinear control via approximate input-output linearization: The ball and beam example, *IEEE Trans. Aut. Con.*, AC-37, 392–398.
- Haykin, S. (1991). *Adaptive Filter Theory*. Prentice-Hall, Englewood Cliffs.
- Haykin, S. (1994). *Neural Networks: A Comprehensive Foundation*. Macmillan.
- Haykin, S. (ed.) (2001). *Kalman Filtering and Neural Networks*. Wiley, New York.
- Hawking, S.W., Ellis, G.F.R. (1973). *The large scale structure of space-time*. Cambridge Univ. Press, London.
- Hawking, S.W. (1975). Particle creation by black holes, *Commun. Math. Phys.* **43**, 199–220.
- Hawking, S.W., Israel, W. (ed.) (1979). *General relativity: an Einstein centenary survey*. Cambridge Univ. Press, Cambridge.
- Hawking, S.W., Penrose, R. (1996). *The Nature of Space and Time*. Princeton Univ. Press, Princeton, NJ.
- Hebb, D.O. (1949). *The Organization of Behaviour*. Wiley, New York.

- Hecht-Nielsen, R. (1987). Counterpropagation networks. *Applied Optics*, **26**(23), 4979–4984.
- Hecht-Nielsen, R. (1990). *NeuroComputing*. Addison–Wesley, Reading.
- Heckhausen, H. (1977). Achievement motivation and its constructs: a cognitive model. *Motiv. Emot.*, **1**, 283–329.
- Heermann, D.W. (1990). *Computer Simulation Methods in Theoretical Physics*. (2nd ed), Springer, Berlin.
- Hehl, F., McCrea, J., Mielke, E., Ne’eman, Y. (1995). Metric–affine gauge theory of gravity. *Phys. Rep.* **258**, 1.
- Helgason, S. (2001). *Differential Geometry, Lie Groups and Symmetric Spaces*. (2nd ed.) American Mathematical Society, Providence, RI.
- Hemami, H., Wyman, B.F. (1979). Modelling and control of constrained dynamic systems with application to biped locomotion in the frontal plane, *IEEE Trans. Aut. Con.*, AC-24, **4**, 526–535.
- Henneaux, M., Teitelboim, C. (1992). *Quantization of Gauge systems*. Princeton Univ. Press, Princeton, NJ.
- Hill, A.V. (1938). The heat of shortening and the dynamic constants of muscle, *Proc. R. Soc. B*, **76**, 136–195.
- Hilbert, D., Cohn-Vossen, S. (1999). *Geometry and the Imagination*. (Reprint ed.), Amer. Math. Soc.
- Hirsch, M.W., Smale, S. (1974). *Differential Equations, Dynamical Systems and Linear Algebra*. Academic Press, New York.
- Hirsch, M.W. (1976). *Differential Topology*. Springer, New York.
- Hirzebruch, F. (1966). *Topological Methods in Algebraic Geometry*, Springer-Verlag, Berlin.
- Hirzebruch, F., Hopf, H. (1958). Felder von Flächenelementen in 4–dimensionalen Mannigfaltigkeiten, *Math. Annalen* **136**, 156;
- Hitchin, N.J. (1997). The moduli space of Special Lagrangian submanifolds. *Annali della Scuola Normale Superiore di Pisa*, **25**(3-4), 503–515.
- Hodgkin, A.L., Huxley, A.F. (1952). A quantitative description of membrane current and application to conduction and excitation in nerve. *J. Physiol.*, **117**, 500–544.
- Hodgkin, A.L. (1964). *The Conduction of the Nervous Impulse*. Liverpool Univ. Press, Liverpool.
- Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational activity. *Proc. Natl. Acad. Sci. USA.*, **79**, 2554–2558.
- Hopfield, J.J. (1984). Neurons with graded response have collective computational properties like those of two–state neurons. *Proc. Natl. Acad. Sci. USA*, **81**, 3088–3092.
- Hopfield, J.J., Tank, D.W. (1985). Neural computation of decisions in optimization problems. *Biol. Cybern.*, **52**, 114–152.
- Hori, K., Katz, S., Klemm, A., Pandharipande, R., Thomas, R., Vafa, C., Vakil, R., Zaslow, E. (2003). *Mirror symmetry*. Clay Mathematics Monographs 1, American Mathematical Society, Providence, Clay Mathematics Institute, Cambridge, MA.
- Horowitz, G.T., Lykken, J., Rohm, R., Strominger, A. (1986). *Purely Cubic*

- Action for String Field Theory, *Phys. Rev. Lett.* **57**, 283.
- Horowitz, G.T. (1989). Exactly soluble diffeomorphism invariant theories. *Comm. Math. Phys.* **125**:3, 417.
- Hoppensteadt, F.C., Izhikevich, E.M. (1997). *Weakly Connected Neural Networks*. Springer, New York.
- Horava, P. (1999). Type IIA D-Branes, K-Theory, and Matrix Theory, *ATMP* **2**, 1373.
- Horsthemke, W., Lefever, R. (1984). *Noise-Induced Transitions*. Springer, Berlin.
- Houk, J.C. (1979). Regulation of stiffness by skeletomotor reflexes. *Ann. Rev. Physiol.*, **41**, 99–114.
- Houk, J.C., Buckingham, J.T., Barto, A.G. (1996). Models of the cerebellum and motor learning. *Behavioral and Brain Sciences*, **19**(3), 368–383.
- Hull, J.C. (2000). *Options, Futures, and Other Derivatives*. (4th ed.), Prentice Hall, New Jersey.
- Humphrey, D.R., Freund, H.J. (1991). *Motor Control: Concepts and Issues*. Wiley, Chichester, UK.
- Hunt, L, Turi, J. (1993). A new algorithm for constructing approximate transformations for nonlinear systems. *IEEE Trans. Aut. Con.*, AC-38, 1553–1556.
- Hurmuzlu, Y. (1993). Dynamics of bipedal gait. *J. Appl. Mech.*, **60**, 331–343.
- Huxley, A.F. (1957). Muscle structure and theories of contraction. *Progr. Biophys. Chem.*, **7**, 255–328.
- Huxley, A.F., Niedergerke, R. (1954). Changes in the cross-striations of muscle during contraction and stretch and their structural interpretation. *Nature*, **173**, 973–976.
- Ichinomiya, T. (2004). Frequency synchronization in random oscillator network. *Phys. Rev. E* **70**, 026116.
- Ichinomiya, T. (2005). Path-integral approach to the dynamics in sparse random network. *Phys. Rev. E* **72**, 016109.
- Igarashi, E., Nogai, T. (1992). Study of lower level adaptive walking in the sagittal plane by a biped locomotion robot. *Advanced Robotics*, **6**, 441–459.
- Ikeda, S. (1990). Some remarks on the Lagrangian Theory of Electromagnetism. *Tensor, N.S.* **49**.
- Immirzi, G. (1997). Quantum Gravity and Regge Calculus. *Nucl. Phys. Proc. Suppl.* **57**, 65–72.
- Ilachinski, A. (2001). *Cellular automata*. World Scientific, Singapore.
- Imamizu, H., Miyauchi, S., Tamada, T., Sasaki, Y., Takino, R., Puetz, B., Yoshioka, T., Kawato, M. (2000). Human cerebellar activity reflecting an acquired internal model of a novel tool. *Nature*, **403**, 192–195.
- Imamizu, H., Kuroda, T., Miyauchi, S., Yoshioka, T., Kawato, M. (2003). Modular organization of internal models of tools in the human cerebellum. *Proc Natl Acad Sci USA*, **100**, 5461–5466.
- Ingber, L. (1997). Statistical mechanics of neocortical interactions: Applications of canonical momenta indicators to electroencephalography, *Phys. Rev. E*, **55**(4), 4578–4593.
- Ingber, L. (1998). Statistical mechanics of neocortical interactions: Training and testing canonical momenta indicators of EEG, *Mathl. Computer Modelling*

27(3), 33–64.

- Inozemtsev, I. (1989). Lax representation with spectral parameter on a torus for integrable particle systems, *Lett. Math. Phys.* **17**, 11–17.
- Inozemtsev, I. (1989). The finite Toda lattices, *Comm. Math. Phys.* **121**, 628–638.
- Isham, C.J., (1997). Structural issues in quantum gravity, in *General Relativity and Gravitation: World Scientific*, Singapore.
- Isham, C.J. (1984). In B. DeWitt, R. Stora, (ed.), *Relativity, Groups and Topology*, Les Houches Session XL. North Holland, Amsterdam.
- Ishibashi, N., Kawai, H., Kitazawa, Y., Tsuchiya, A. (1997). A Large-N Reduced Model as Superstring. *Nucl. Phys.* **B498**, 467.
- Isidori, A. (1989). *Nonlinear Control Systems. An Introduction*, (2nd ed) Springer, Berlin.
- Ito, K. (1960). Wiener Integral and Feynman Integral. *Proc. Fourth Berkeley Symp. Math., Stat., Prob.*, **2**, 227–238.
- Ivancevic, V. (1991). *Introduction to Biomechanical Systems: Modelling, Control and Learning* (in Serbian). Scientific Book, Belgrade.
- Ivancevic, V., Lukman, L., Ivancevic, T. (1995) *Selected Chapters in Human Biomechanics. Textbook* (in Serbian). Univ. Novi Sad Press, Novi Sad.
- Ivancevic, T. (1995). Some Possibilities of Multilayered Neural Networks' Application in Biomechanics of Muscle Contractions, Man Locomotion and Sport Training, Master Thesis (in Serbo-Croatian), Univ. Novi Sad.
- Ivancevic, T., Jain, L.C., Bottema, M. (1999). New Two-feature GBAM–Neurodynamical Classifier for Breast Cancer Diagnosis, *Proc. from KES'99*, IEEE Press.
- Ivancevic, T., Jain, L.C., Bottema, M. (1999). A New Two-Feature FAM–Matrix Classifier for Breast Cancer Diagnosis, *Proc. from KES'99*, IEEE Press.
- Ivancevic, V., Snoswell, M. (2000). *Torson Sports Cybernetics*. Torson Group. Inc., Adelaide.
- Ivancevic, V., Snoswell, M. (2001). Fuzzy–stochastic functor machine for general humanoid–robot dynamics. *IEEE Trans. on Sys, Man, Cyber. B*, **31**(3), 319–330.
- Ivancevic, V., Pearce, C.E.M. (2001). Poisson manifolds in generalised Hamiltonian biomechanics. *Bull. Austral. Math. Soc.* **64**, 515–526.
- Ivancevic, V., Pearce, C.E.M. (2001). Topological duality in humanoid robot dynamics. *ANZIAM J.* **43**, 183–194.
- Ivancevic, V. (2002). Generalized Hamiltonian biodynamics and topology invariants of humanoid robots. *Int. J. Mat. Mat. Sci.*, **31**(9), 555–565.
- Ivancevic, V. (2004). Symplectic Rotational Geometry in Human Biomechanics. *SIAM Rev.*, **46**(3), 455–474.
- Ivancevic, V. (2005). Dynamics of Humanoid Robots: Geometrical and Topological Duality. To appear in *Biomathematics: Modelling and simulation* (ed. J.C. Misra), World Scientific, Singapore.
- Ivancevic, V. (2005). A Lagrangian model in human biomechanics and its Hodge–de Rham cohomology. *Int. J. Appl. Math. Mech.*, **3**, 35–51.
- Ivancevic, V. (2005). Lie–Lagrangian model for realistic human bio-dynamics. *Int. J. Humanoid Robotics* (World Sci.) (to appear)

- Ivancevic, V., Beagley, N. (2005). Brain-like functor control machine for general humanoid biodynamics. *Int. J. Math. Math. Sci.* **11**, 1759–1779.
- Ivancevic, V., Ivancevic, T. (2006). *Natural Biodynamics*. World Scientific, Singapore.
- Ivancevic, V., Pearce, C.E.M. (2006). Hamiltonian dynamics and Morse topology of humanoid robots. *Gl. J. Mat. Math. Sci.* (to appear)
- Iyanaga, S., Kawada, Y. (eds.) (1980). Pontrjagin's Maximum Principle. In *Encyclopedic Dictionary of Mathematics*. MIT Press, Cambridge, MA, 295–296.
- Jibu, M., Yasue, K. (1995). *Quantum brain dynamics and consciousness*. John Benjamins, Amsterdam.
- Jibu, M., Pribram, K.H., Yasue, K. (1996). From conscious experience to memory storage and retrieval: the role of quantum brain dynamics and boson condensation of evanescent photons, *Int. J. Mod. Phys. B*, **10**, 1735.
- Jones, V.F.R. (1985). A new polynomial invariant of knots and links, *Bull. Amer. Math. Soc.* **12**, 103.
- Jost, J., Peng, X., Wang, G. (1995). Variational Aspects of The Seiberg-Witten Functional, arXiv:dg-ga 9504003.
- Kac, M. (1951). On Some Connection between Probability Theory and Differential and Integral Equations. *Proc. 2nd Berkeley Sympos. Math. Stat. and Prob.*, 189–215.
- Kadic, A., Edelen, D.G.B. (1983). *A Gauge theory of Dislocations and Disclinations*, New York, Springer.
- Kaku, M. (1988). *Introduction to superstrings*. Springer-Verlag.
- Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME, Ser. D, J. Bas. Eng.*, **82**, 34–45.
- Kalman, R.E., Falb, P., Arbib, M.A. (1969). *Topics in Mathematical System Theory*. McGraw Hill, New York.
- Kamber, F., Tondeur, P. (1975). *Foliated Bundles and Characteristic Classes*. *Lecture Notes in Math.* 493, Springer, Berlin.
- Kamron, N., Olver, P.J. (1989). Le problème d'équivalence à une divergence prés dans le calcul des variations des intégrales multiples. *C. R. Acad. Sci. Paris*, t. **308**, Série I, 249–252.
- Kan, D.M. (1958). Adjoint Functors. *Trans. Am. Math. Soc.* **89**, 294–329.
- Kanatchikov, I. (1998). Canonical structure of classical field theory in the polynomial phase space. *Rep. Math. Physics*, **41**(1), 49–90.
- Kappen, H.J. (2006). A linear theory for control of nonlinear stochastic systems. *Phys. Rev. Let.* (to appear)
- Kasabov, N. (2002). *Evolving connectionist systems: Methods and applications in bioinformatics, brain study and intelligent machines*. Springer, London.
- Katok, A. (1980). Lyapunov exponents, entropy and periodic orbits for diffeomorphisms. *Inst. Hautes Études Sci. Publ. Math.* **51**, 137–173.
- Katz, S., Klemm, A., Vafa, C. (1999). M-theory, topological strings and spinning black holes. *Adv. Theor. Math. Phys.* **3**, 1445.
- Kauffman, L.H. (1994). *Knots and Physics* (2nd ed). World Scientific, Singapore.
- Kauffman, L.H., Radford, D.E. (1995). Invariants of 3-manifolds derived from finite-dimensional Hopf algebras. *J. Knot Theory Ramifications* **4**, 131–162.

- Kauffman, L.H., Radford, D.E. (1999). Quantum algebra structures on $n \times n$ matrices. *J. Algebra* **213**, 405–436.
- Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, **9**, 718–727.
- Kedra, J. (2000). On the Flux group. *Math. Res. Lett.* **7**, 279–285.
- Kelly, S.D., Murray, R.M. (1995). Geometric phases and robotic locomotion. *J. Robotic Systems*, **12**(6), 417–431.
- Kijowski, J., Szczyrba, W. (1976). A Canonical Structure for Classical Field Theories. *Commun. Math. Phys.* **46**, 183.
- Kijowski, J., Tulczyjew, W. (1979). *A Symplectic Framework for Field Theories*. Springer-Verlag, Berlin.
- Klauder, J.R. (1997). Understanding Quantization. *Found. Phys.* **27**, 1467–1483.
- Klauder, J.R. (2000). *Beyond Conventional Quantization*, Cambridge Univ. Press, Cambridge, UK.
- Klimyk, A., Schmüdgen, K. (1997). *Quantum groups and their representations*, Springer, Berlin.
- Kobayashi, S., Nomizu, K. (1963/9). *Foundations of Differential Geometry*, Vols. 1,2., Interscience Publ., New York.
- Kock, A., (1981). *Synthetic Differential Geometry*, London Math.Soc. Lecture Notes Series No. 51, Cambridge Univ. Press, Cambridge.
- Kock, A. (2001). Infinitesimal aspects of the Laplace operator. *Theory and Applications of Categories*, **9**(1), 1–16.
- Kock, A., Reyes, G.E. (2003). Some calculus with extensive quantities: wave equation. *Theory and Applications of Categories*, **11**(14), 321–336.
- Kohonen, T. (1988). *Self Organization and Associative Memory*. Springer, Berlin.
- Kolar, I., Michor, P.W., Slovák, J. (1993). *Natural Operations in Differential Geometry*. Springer, Berlin.
- Koon, W.S., Marsden, J.E. (1997). The Hamiltonian and Lagrangian approaches to the dynamics of nonholonomic systems. *Reports on Math Phys.* **40**, 21–62.
- Kosko, B. (1992). *Neural Networks and Fuzzy Systems, A Dynamical Systems Approach to Machine Intelligence*. Prentice-Hall, New York.
- Kosko, B. (1993). *Fuzzy Thinking*. Disney Books, Hyperion.
- Kosko, B. (1996). *Fuzzy Engineering*. Prentice Hall, New York.
- Kosko, B. (1999). *The Fuzzy Future: From Society and Science to Heaven in a Chip*. Random House, Harmony.
- Kosmann-Schwarzbach, Y. (2004). Derived brackets. *Lett. Math. Phys.* **69**, 61–87.
- Kotschik, D., Morgan, J.W., Taubes, C.H. (1995). Four manifolds without symplectic structures but with non-trivial Seiberg-Witten invariants, *Math. Research Letters* **2**, 119–124.
- Kotz, S., Nadarajah, S. (2000). *Extreme Value Distributions*. Imperial College Press, London.
- Krasil'shchik, I., Lychagin, V., Vinogradov, A. (1985). *Geometry of Jet Spaces and Nonlinear Partial Differential Equations*, Gordon and Breach, Glasgow.
- Krasnov, K. (1997). Geometrical entropy from loop quantum gravity, *Phys. Rev.* **D55**, 3505.

- Krener, A. (1984). Approximate linearization by state feedback and coordinate change, *Systems Control Lett.*, 5, 181–185.
- Kronheimer, P., Mrowka, T. (1994). The genus of embedded surfaces in the projective plane, *Math. Research Letters* **1**, 797–808;
- Kronheimer, P., Mrowka, T. (1994). Recurrence relations and asymptotics for four-manifold invariants, *Bull. Amer. Math. Soc.* **30**, 215.
- Krupkova, O. (1997). *The Geometry of Ordinary Variational Equations*. Springer, Berlin.
- Kuhl, J. (1985). Volitional Mediator of cognition-Behaviour consistency: Self-regulatory Processes and action versus state orientation (pp. 101–122). In: J. Kuhl & S. Beckman (Eds.) *Action control: From Cognition to Behaviour*. Springer, Berlin.
- Kuramoto, Y. (1984). *Chemical Oscillations. Waves and Turbulence*, Springer, New York.
- Labastida, J.M.F., Lozano, C. (1998). Lectures on topological quantum field theory. In *Proceedings of the CERN–Santiago de Compostela–La Plata Meeting on ‘Trends in Theoretical Physics’*, eds. H. Falomir, R. Gamboa, F. Schaposnik, Amer. Inst. Physics, New York.
- Lafferriere, G., Sussmann, J.J. (1993). A differential geometric approach to motion planning. In Z. Li and J. F. Canny (ed.) *Nonholonomic Motion Planning*, 235–270. Kluwer, Dordrecht.
- Lalonde, F., McDuff, D. (2002). Symplectic structures on fiber bundles. *Topology* **42**(2), 309–347.
- Lalonde, F., McDuff, D., Polterovich, L. (1998). On the Flux conjectures, *CRM Proceedings and Lecture Notes* vol **15**, 69–85.
- Lalonde, F., McDuff, D., Polterovich, L. (1999). Topological rigidity of Hamiltonian loops and quantum homology. *Invent. Math.* **135**, 369–385.
- Lanczos, C. (1986). *The variational principles of mechanics*. Dover, New York.
- Landau, L.D., Lifshitz, E.M. (1977). *Quantum Mechanics: Non-Relativistic Theory*, Pergamon Press, Oxford.
- Landau, L.D., Lifshitz, E.M. (1978). *Statistical Physics*, Pergamon Press, Oxford.
- Landi, G. (1998). *An introduction to Noncommutative Spaces and Their Geometries*, Springer, Berlin.
- Landsman, N.P. (1995). Against the Wheeler–DeWitt equation. *Class. Quant. Grav.* **12**, 119–124.
- Lang, S. (1999). *Fundamentals of Differential Geometry*. Graduate Texts in Mathematics, Springer, New York.
- Lang, S. (2002). *Introduction to Differentiable Manifolds* (2nd ed.). Graduate Texts in Mathematics, Springer, New York.
- Lang, S. (2003). *Complex Analysis* (rev. 4th ed). Graduate Texts in Mathematics, Springer, New York.
- Lang, S. (2005). *Algebra* (rev. 3rd ed.). Graduate Texts in Mathematics, Springer, New York.
- Langer, J., Perline, R. (1994). Local geometric invariants of integrable evolution equations, *J. Math. Phys.*, **35**(4), 1732–1737.
- Langouche, F., Roekaerts, D., Tirapegui, E. (1980). Short derivation of Feynman

- Lagrangian for general diffusion process. *J. Phys. A* **113**, 449-452.
- Lax, P.D. (1968). Integrals of nonlinear equations of evolution and solitary waves. *Comm. Pure Appl. Math.*, **21**, 467-490.
- Lawvere, F.W. (1979). Categorical Dynamics. *Topos Theoretic Methods in Geometry* (ed. A. Kock). Aarhus Various Publ, Series **30**.
- Lawson, H.B., Michelsohn, M.L. (1989). *Spin Geometry*. Princeton Univ. Press, Princeton, NJ.
- LeBrun, C. (1991). Scalar-Flat Kähler Metrics On Blown-Up Ruled Surfaces. *J. Reine Angew Math.* **420**, 161.
- de León, M., Marrero, J. (1993). Constrained time-dependent Lagrangian systems and Lagrangian submanifolds, *J. Math. Phys.* **34**, 622.
- de León, M., Martín, D., de Diego, A., Santamaria, M. (2004). Symmetries in Classical Field Theory. [arXiv:math-ph/0404013](https://arxiv.org/abs/math-ph/0404013).
- de León, M., Martín, D., de Diego, A. (1996). On the geometry of nonholonomic Lagrangian systems. *J. Math. Phys.* **37**, 3389.
- de León, M., Marrero, J., Martín, D., de Diego D. (1997). Nonholonomic Lagrangian systems in jet manifolds. *J. Phys. A*, **30**, 1167.
- Leinster, T. (2002). A survey of definitions of n -category. *Theor. Appl. Categ.* **10**, 1-70.
- Leinster, T. (2003). *Higher Operads, Higher Categories*, London Mathematical Society Lecture Notes Series, Cambridge Univ. Press, Cambridge.
- Leinster, T. (2004). Operads in higher-dimensional category theory. *Theor. Appl. Categ.* **12**, 73-194.
- Lerche, W., Vafa, C., Warner, N.P. (1989). Chiral Rings. In *N=2 Superconformal Theories*, *Nucl. Phys. B* **324**, 427.
- Lewin, K. (1951). *Field Theory in Social Science*. Univ. Chicago Press, Chicago.
- Lewin, K. (1997). *Resolving Social Conflicts: Field Theory in Social Science*, American Psych. Assoc., New York.
- Lewis, A.D., Murray, R.M. (1997). Controllability of simple mechanical control systems, *SIAM J. Con. Opt.*, **35**(3), 766-790.
- Lewis, A.D. (1995). *Aspects of Geometric Mechanics and Control of Mechanical Systems*. Technical Report CIT-CDS 95-017 for the Control and Dynamical Systems Option, California Institute of Technology, Pasadena, CA.
- Lewis, A.D., Murray, R.M. (1999). Configuration controllability of simple mechanical control systems, *SIAM Review*, **41**(3), 555-574.
- Lewis, A.D. (1998). Affine connections and distributions with applications to nonholonomic mechanics, *Reports on Mathematical Physics*, **42**(1/2), 135-164.
- Lewis, A.D. (1999). When is a mechanical control system kinematic? In *Proceedings of the 38th IEEE Conf. Decis. Con.*, 1162-1167, IEEE, Phoenix, AZ.
- Lewis, A.D. (2000). Simple mechanical control systems with constraints. *IEEE Trans. Aut. Con.*, **45**(8), 1420-1436.
- Lewis, A.D. (2000). Affine connection control systems. *Proceedings of the IFAC Workshop on Lagrangian and Hamiltonian Methods for Nonlinear Control* 128-133, Princeton.

- Li, M., Vitanyi, P. (1997). *An Introduction to Kolmogorov Complexity and its Applications*. Springer, New York.
- Liebermann, P., Marle, C.M. (1987). *Symplectic Geometry and Analytical Mechanics*, Reidel, Dordrecht.
- Lieh, J. (1994). Computer oriented closed-form algorithm for constrained multi-body dynamics for robotics applications. *Mechanism and Machine Theory*, **29**, 357–371.
- Linz, S.J. (1997). Nonlinear Dynamical Models and Jerky Motion. *Am. J. Phys.*, **65**(6), 523–526.
- Liu, G., Tian, G. (1998). Floer homology and Arnold conjecture, *J. Diff. Geom.*, **49**, 1–74.
- Loll, R. (1998). Discrete approaches to quantum gravity in four dimensions. *Living Reviews in Relativity*, **13**.
- Loll, R. (2001). Discrete Lorentzian quantum gravity. *Nucl. Phys. B* **94**, 96–107.
- Loo, K. (1999). A Rigorous Real Time Feynman Path Integral. *J. Math. Phys.*, **40**(1), 64–70.
- Loo, K. (2000). A Rigorous Real Time Feynman Path Integral and Propagator. *J. Phys. A: Math. Gen.*, **33**, 9215–9239.
- Lopez, M.C., Marsden, J.E. (2003). Some remarks on Lagrangian and Poisson reduction for field theories. *J. Geom. Phys.*, **48**, 52–83.
- Lorenz, E.N., (1963). Deterministic Nonperiodic Flow. *J. Atmos. Sci.*, **20**, 130–141.
- Louie, A.H. (1983). Categorical system theory and the phenomenological calculus. *Bull. Math. Biol.*, **45**, 1029–1045; Categorical system theory. *Bull. Math. Biol.*, **45**, 1047–1072.
- Louko, J., Sorkin, R.D. (1997). Complex actions in two-dimensional topology change. *Class. Quant. Grav.* **14**, 179–204.
- Lu, J.-H. (1990). Multiplicative and affine Poisson structures on Lie groups. PhD Thesis, Berkeley Univ., Berkeley.
- Lu, J.-H. (1991). Momentum mappings and reduction of Poisson actions. In *Symplectic Geometry, Groupoids, and Integrable Systems*, eds.: P. Dazord and A. Weinstein, 209–225, Springer, New York.
- Lygeros, J., Godbole, D.N., Sastry, S. (1998). Verified hybrid controllers for automated vehicles, *IEEE Trans. Aut. Con.*, **43**, 522–539.
- MacLane, S. (1971). *Categories for the Working Mathematician*. Springer, New York.
- Madore, J. (1995). *An introduction to Noncommutative Differential Geometry*, LMS Lecture Notes **206**.
- Mangiarotti, L., Sardanashvily, G. (1998). *Gauge Mechanics*. World Scientific, Singapore.
- Mangiarotti, L., Sardanashvily, G. (1999). On the geodesic form of non-relativistic dynamic equations. arXiv:math-ph/9906001.
- Mangiarotti, L., Sardanashvily, G. (2000). *Connections in Classical and Quantum Field Theory*, World Scientific, Singapore.
- Mangiarotti, L., Sardanashvily, G. (2000). Constraints in Hamiltonian time-dependent mechanics. *J. Math. Phys.* **41**, 2858.

- Mangiarotti, L., Obukhov, Yu., Sardanashvily, G. (1999). *Connections in Classical and Quantum Field Theory*. World Scientific, Singapore.
- Mangioni, S.E., Deza, R.R., Toral, R., Wio, H.S. (2000). Nonequilibrium phase transitions induced by multiplicative noise: effects of self-correlation. *Phys. Rev. E*, **61**, 223–231.
- Manikonda, V. (1998). *Control and Stabilization of a Class of Nonlinear Systems with Symmetry*. PhD Thesis, Center for Dynamics and Control of Smart Structures, Harvard Univ., Cambridge.
- Marathe, K., Martucci, G. (1992). *The Mathematical Foundations of Gauge Theories*. North-Holland, Amsterdam.
- Marculli, M. (1995). Notes on Seiberg-Witten Gauge Theory. arXiv:dg-ga/9509005.
- Marmo, G., Simoni, A., Stern, A. (1995). Poisson Lie group symmetries for the isotropic rotator. *Int. J. Mod. Phys. A* **10**, 99–114.
- Marsden, J.E., Ratiu, T.S. (1999). *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*. (2nd ed), Springer, New York.
- Marsden, J.E., Weinstein, A. (1974). Reduction of symplectic manifolds with symmetry. *Rept. Math. Phys.* **5**, 121–130.
- Marsden, J.E., Patrick, G.W., Shkoller, S. (1998). Multisymplectic Geometry, Variational Integrators, and Nonlinear PDEs. *Comm. Math. Phys.*, **199**, 351–395.
- Marshakov, A. (1997). On integrable systems and supersymmetric gauge theories, *Theor. Math. Phys.* **112**, 791–826.
- Martini, R., Gragert, P.K.H. (1999). Solutions of WDVV equations in Seiberg-Witten theory from root systems. *J. Nonlin. Math. Phys.* **6**:1, 1–4.
- Mascalchi, M. *et al.* (2002). Proton MR Spectroscopy of the Cerebellum and Pons in Patients with Degenerative Ataxia, *Radiology*, **223**, 371.
- Massa, E., Pagani, E. (1994). Jet bundle geometry, dynamical connections and the inverse problem of Lagrangian mechanics. *Ann. Inst. Henri Poincaré* **61**, 17.
- Massey, W.S. (1967). *Algebraic Topology: An Introduction*, Harcourt, Brace & World.
- Matacz, A. (2002). Path Dependent Option Pricing: the Path Integral Partial Averaging Method. *J. Comp. Finance*, **6**(2).
- Mathai V., Quillen, D. (1986). Super-connections, Thom classes and equivariant differential forms, *Topology* **25**, 85–110.
- Matsumoto, M. (1982). *Foundations of Finsler Geometry and Special Finsler Spaces*. Kaisheisha Press, Kyoto.
- Mavromatos, N.E., Nanopoulos, D.V. (1995). A Non-critical String (Liouville) Approach to Brain Microtubules: State Vector reduction, Memory coding and Capacity. ACT-19/95, CTP-TAMU-55/95, OUP-95-52P.
- Mavromatos, N.E., Nanopoulos, D.V. (1995). Non-Critical String Theory Formulation of Microtubule Dynamics and Quantum Aspects of Brain Function. ENSLAPP-A-524/95.
- May, R. (1976). *Simple Mathematical Models with Very Complicated Dynamics*.

- Nature, **261**(5560), 459-467.
- Maybeck, P.S. (1979). *Stochastic Models, Estimation, and Control*, Volume 1, Academic Press, New York.
- Mayer, P.A. (1981). *A Differential Geometric Formalism for the Ito Calculus*. Lecture Notes in Mathematics, **851**, Springer, New York.
- McCullagh, P. (1987). *Tensor methods in statistics*. Monographs in statistics and applied probability. Chapman & Hall, Cambridge, UK.
- McDuff, D. (1988). The moment map for circle actions on symplectic manifolds. *J. Geom. Phys.* **5**, 149-60.
- McDuff, D. (1999). The virtual moduli cycle. *Amer. Math. Soc. Transl.* **2**(196), 73-102.
- McDuff, D. (2000). Quantum homology of fibrations over S^2 , *Int. Math. J.* **11**, 665-721.
- McDuff, D., Salamon, D.A. (1998). *Introduction to Symplectic Topology* (2nd ed), OUP, Oxford.
- McGeer, T. (1990). Passive dynamic walking. *Int. J. Robot. Res.*, **9**, 62-82.
- Merton, R. (1973). *Theory of Rational Option Pricing*. *Bell J. Econom. Managem. Sci.* **4**, 141-183.
- Messiah, A. (2000). *Quantum Mechanics: Two Volumes Bound As One*. Dover Pubs.
- Metzger, M.A. (1997). Applications of nonlinear dynamical systems theory in developmental psychology: Motor and cognitive development. *Nonlinear Dynamics, Psychology, and Life Sciences*, **1**, 55-68.
- Meyer, K.R., Hall, G.R. (1992). *Introduction to Hamiltonian Dynamical Systems and the N-body Problem*. Springer, New York.
- Meyer, K.R., Schmidt, D.S. (2000). From the restricted to the full three-body problem. *Trans. Amer. Math. Soc.*, **352**, 2283-2299.
- Meyer, F. (2005). *Noncommutative Spaces and Gravity*. Proc. I Modave Summer School in Math. Phys. Modave, Belgium.
- Michor, P.W. (2001). *Topics in Differential Geometry*. Lecture notes of a course in Vienna.
- Milinkovic, D. (1999). Morse homology for generating functions of Lagrangian submanifolds. *Trans. Amer. Math. Soc.* **351**(10), 3953-3974.
- Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information, *Psych. Rev.*, **63**, 81-97.
- Milnor, J. (1963). *Morse Theory*. Princeton Univ. Press, Princeton, NJ.
- Milnor, J. (1965). *Lectures on the H-Cobordism Theorem*. Math. Notes. Princeton Univ. Press, Princeton.
- Milnor, J.W., Stasheff, J.D., (1974). *Characteristic Classes*, Princeton Univ. Press, Princeton, NJ.
- Milnor, J. (1997). *Topology from the Differentiable Viewpoint* (rev. ed.). Princeton Univ. Press, Princeton.
- Miron, R., Anastasiei, M. (1994). *The Geometry of Lagrangian Spaces: Theory and Applications*, Kluwer Academic Publishers.
- Miron, R., Kirkovits, M.S., Anastasiei, M. (1988). A Geometrical Model for Variational Problems of Multiple Integrals. *Proc. Conf. Diff. Geom Appl.*

- Dubrovnik, Yugoslavia.
- Mironov, A. (1998). WDVV Equations in Seiberg-Witten theory and associative algebras. *Nucl.Phys.Proc.Suppl.* **61A**, 177-185.
- Misner, C.W., Thorne, K.S., Wheeler, J.A. (1973). *Gravitation*. Freeman, San Francisco.
- Mladenova, C. (1991). Mathematical Modelling and Control of Manipulator Systems. *Int. J. Rob. Comp. Int. Man.*, **8**(4), 233-242.
- Mladenova, C. (1999). Applications of Lie Group Theory to the Modelling and Control of Multibody Systems. *Mult. Sys. Dyn.*, **3**(4), 367-380.
- Modugno, M., Stefani, G. (1978). Some results on second tangent and cotangent spaces. *Quadernidell' Instituto di Matematica dell' Universit a di Lecce Q.*, **16**.
- Moerdijk, I., Reyes, G.E. (1991). *Models for Smooth Infinitesimal Analysis*. Springer, New York.
- Montagna, G., Nicosini, O., Moreni, N. (2002). A path integral way to option pricing. *Physica A* **310**, 450-466.
- C. Montonen, D. Olive (1977). Magnetic Monopoles as Gauge Particles? *Phys. Lett.* **B72**, 117.
- Morandi, G., Ferrario, C., Lo Vecchio, G., Marmo, G., Rubano, C. (1990). The inverse problem in the calculus of variations and the geometry of the tangent bundle, *Phys. Rep.* **188**, 147.
- Moroianu, A., Semmelmann, U. (2003). Twistor forms on Kähler manifolds. *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* **2**(4), 823-845, (2003)
- Morse, M. (1934). *The Calculus of Variations in the Large*. Amer. Math. Soc. Coll. Publ. No. 18, Providence, RI.
- Mosekilde, E. (1996). *Topics in Nonlinear Dynamics: Application to Physics, Biology and Economics*. World Scientific, Singapore.
- Moser, J. (1975). Three integrable Hamiltonian systems connected with isospectral deformations, *Advances Math.* **16**, 197.
- Moyal, J.E. (1949). Quantum mechanics as a statistical theory. *Proc. Cambridge Phil. Soc.* **45**, 99-124.
- Muñoz-Lecanda, M. and Román-Roy, N. (1992). Lagrangian theory for presymplectic systems. *Ann. Inst. Henri Poincaré* **57**, 27.
- Munkres, J. (1999). *Topology* (2nd ed.). Prentice Hall.
- Munteanu, F., Rey, A.M., Salgado, M. (2004). The Günther's formalism in classical field theory: momentum map and reduction. *J. Math. Phys.* **45**(5), 1730-1750.
- Murray, J.D. (1989). *Mathematical Biology*. Springer, Berlin.
- Murray, R.M., Li, X., Sastry, S. (1994). *Robotic Manipulation*, CRC Press, Boca Raton, FL.
- Murray, R.M. (1994). Nilpotent bases for a class of nonintegrable distributions with applications to trajectory generation for nonholonomic systems. *Math. Control Signals Systems*, **7**, 58-75.
- Mustafa, M.T. (1999). Restrictions on harmonic morphisms. *Confomal Geometry and Dynamics (AMS)*, **3**, 102-115.
- Myung, I.J., Pitt, M.A. (2003). Model Evaluation, Testing and Selection. In K.

- Lambert and R. Goldstone (eds.), *Handbook of Cognition*, Sage Publ., (2003)
- Myung, I.J., Balasubramanian, V., Pitt, M.A. (2000). Counting probability distributions: differential geometry and model selection. *Proceedings of the National Academy of Science, USA*, **97**, 11170–11175.
- Myung, I.J., Forster, M., Browne, M.W. (2000). Special issue on model selection. *J. Math. Psych.*, **44**, 1–2.
- Nakatsu, T., Takasaki, K. (1996). Whitham-Toda Hierarchy and $N = 2$ supersymmetric Yang–Mills Theory, *Mod. Phys. Lett.* **A11**, 157.
- Nanopoulos, D.V. (1995). *Theory of Brain Function, Quantum Mechanics and Superstrings*. CERN-TH/95128.
- Nash, C., Sen, S. (1983). *Topology and Geometry for Physicists*. Academic Press, London.
- Nayfeh, A.H. (1973). *Perturbation Methods*. Wiley, New York.
- Neagu, M. (2000). Upon h –normal Γ –linear connection on $J^1(T, M)$. arXiv:math.DG/0009070.
- Neagu, M. (2002). The Geometry of Autonomous Metrical Multi-Time Lagrange Space of Electrodynamics. *Int. J. Math. Math. Sci.* **29**, 7-16.
- Neagu, M. Udriște, C. (2000). Multi-Time Dependent Sprays and Harmonic Maps on $J^1(T, M)$, *Third Conference of Balkan Society of Geometers*, Politehnica University of Bucharest, Romania.
- Neagu, M. Udriște, C. (2000). Torsions and Curvatures on Jet Fibre Bundle $J^1(T, M)$. arXiv:math.DG/0009069.
- Nekrasov, N. (1996). Holomorphic bundles and many-body systems, *Comm. Math. Phys.* **180**, 587.
- von Neumann, J. (1955). *Mathematical Foundations of Quantum Mechanics*. Princeton Univ. Press, Princeton.
- Nicolis, G., Prigogine, I. (1977). *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order through Fluctuations*. Wiley Europe.
- Nijmeijer, H., van der Schaft, A.J. (1990). *Nonlinear Dynamical Control Systems*. Springer, New York.
- Nikitin, I.N. (1995). Quantum string theory in the space of states in an indefinite metric. *Theor. Math. Phys.* **107**(2), 589601.
- Nitta, T. (1997). An extension of the back-propagation algorithm to complex numbers. *Neu. Net.*, **10**(8), 1392–1415.
- Nitta, T. (2000). An analysis on fundamental structure of complex-valued neuron. *Neu. Proc. Let.*, **12**(3), 239–246.
- Nitta, T. (2004). Reducibility of the Complex-valued Neural Network. *Neu. Inf. Proc.*, **2**(3), 53–56.
- Noakes, L., Heinzinger, G., Paden, B. (1989). Cubic splines on curved spaces, *IMA J. Math. Con. Inf.*, **6**(4), 465–473.
- Noble, D. (1962). A modification of the Hodgkin–Huxley equations applicable to Purkinie fibre action and peace-maker potentials. *J. Physiol.* **160**, 317–330.
- Novikov, S.P., Manakov, S.V., Pitaevskii, L.P., Zakharov, V.E. (1984). *Theory of Solitons*, Plenum/Kluwer,
- Ohta, Y. (1998). *Topological Field Theories associated with Three Dimensional*

- Seiberg-Witten monopoles. *Int.J.Theor.Phys.* **37**, 925-956.
- Oja, E. (1982). A simplified neuron modeled as a principal component analyzer. *J. Math. Biol.* **15**, 267-273.
- Olshanetsky, M.A., Perelomov, A.M. (1976). Completely integrable Hamiltonian systems connected with semisimple Lie algebras, *Inventiones Math.* **37**, 93-108.
- Olshanetsky, M.A., Perelomov, A.M. (1981). Classical integrable finite-dimensional systems related to Lie algebras, *Phys. Rep.* **71C**, 313-400.
- Olver, P.J. (1986). *Applications of Lie Groups to Differential Equations* (2nd ed.) Graduate Texts in Mathematics, vol. 107, Springer, New York.
- Omohundro, S.M. (1986). *Geometric Perturbation Theory in Physics*. World Scientific, Singapore.
- Ooguri, H. (1992). Partition Functions and Topology-Changing Amplitudes in the 3D Lattice Gravity of Ponzano and Regge. *Nucl. Phys.* **B382**, 276.
- Ooguri, H. (1992). Topological Lattice Models in Four Dimensions. *Mod. Phys. Lett.* **A7**, 2799-2810.
- Ooguri, H., Strominger, A., Vafa, C. (2004). Black hole attractors and the topological string. *Phys. Rev. D* **70**, 106007.
- Ori, A., Piran, T. (1987). Naked singularities in self-similar spherical gravitational collapse. *Phys. Rev. Lett.*, **59**, 2137-2140.
- Oriti, D. (2001). Spacetime geometry from algebra: Spin foam models for non-perturbative quantum gravity. *Rept. Prog. Phys.* **64**, 1489-1544.
- Oseledets, V.I. (1968). A Multiplicative Ergodic Theorem: Characteristic Lyapunov Exponents of Dynamical Systems. *Trans. Moscow Math. Soc.*, **19**, 197-231.
- Ott, E., Grebogi, C., Yorke, J.A. (1990). Controlling chaos. *Phys. Rev. Lett.*, **64**, 1196-1199.
- Ozsvath, I. (1970). Dustfilled universes of class II and III. *J. Math. Phys.*, **11**, 2871-2883.
- Pappas, G.J., Lafferriere, G., Sastry, S. (2000). Hierarchically consistent control systems. *IEEE Trans. Aut. Con.*, **45**(6), 1144-1160.
- Pappas, G.J., Simic, S. (2002). Consistent hierarchies of affine nonlinear systems. *IEEE Trans. Aut. Con.*, **47**(5), 745-756.
- Park, J., Chung, W.-K. (2005). Geometric Integration on Euclidean Group With Application to Articulated Multibody Systems. *IEEE Trans. Rob.* **21**(5), 850-863.
- Park, J.K., Steiglitz, K. and Thurston, W.P., (1986). Soliton-like Behavior in Automata. *Physica D*, **19**, 423-432.
- Paulfler, C., Römer, H. (2002). The geometry of Hamiltonian n -vector-fields in multisymplectic field theory. *J. Geom. Phys.* **44**(1), 52-69.
- Paul, W., Baschnagel, J. (1999). *Stochastic Processes: from Physics to Finance*. Springer, Berlin.
- Pearce, C.E.M., Ivancevic, V. (2003). A generalised Hamiltonian model for the dynamics of human motion. In *Differential Equations and Applications*, Vol. 2, Eds. Y.J. Cho, J.K. Kim and K.S. Ha, Nova Science, New York.
- Pearce, C.E.M., Ivancevic, V. (2004). A qualitative Hamiltonian model for the

- dynamics of human motion. In *Differential Equations and Applications*, Vol. 3, Eds. Y.J. Cho, J.K. Kim and K.S. Ha, Nova Science, New York.
- Pecora, L.M., Carroll T.L. (1991). Driving systems with chaotic signals. *Phys. Rev. A*, **44**, 2374–2383.
- Pecora, L.M., Carroll T.L. (1998). Master stability functions for synchronized coupled systems. *Phys. Rev. Lett.* **80**, 2109–2112.
- Pellionisz, A., Llinas, R. (1979). Brain modeling by tensor network theory and computer simulation. *NeuroScience*, **4**, 323–348.
- Penrose, R. (1967). Twistor algebra. *J. Math. Phys.*, **8**, 345–366.
- Penrose, R. (1971). Angular momentum: an approach to combinatorial space–time, in Bastin, T. (ed.), *Quantum Theory and Beyond*, 151–180, Cambridge Univ. Press, Cambridge, UK.
- Penrose, R. (1971). Applications of negative dimensional tensors, in Welsh, D. (ed.) *Combinatorial Mathematics and its Application*, 221–243.
- Penrose, R. (1989). *The Emperor’s New Mind*, Oxford Univ. Press, Oxford.
- Penrose, R. (1994). *Shadows of the Mind*. Oxford Univ. Press, Oxford.
- Penrose, R. (1997). *The Large, the Small and the Human Mind*. Cambridge Univ. Press, Cambridge.
- Penrose, R., MacCallum, M.A.H. (1972). Twistor theory: an approach to the quantization of fields and space–time, *Phys. Rep.* **6C**, 241–315.
- Penrose, R., Rindler, W. (1984). *Spinors and Space-time*, Vol. I, Cambridge Univ. Press, Cambridge, UK.
- Penrose, R., Rindler, W. (1986). *Spinors and Space-time*, Vol. II, Cambridge Univ. Press, Cambridge, UK.
- Pessa, E., Vitiello, G. (1999). Quantum dissipation and neural net dynamics, *Bioelectrochem. Bioener.*, **48**, 339–342.
- Pessa, E., Vitiello, G. (2003). Quantum noise, entanglement and chaos in the quantum field theory of mind/brain states, *Mind and Matter*, **1**, 59–79.
- Pessa, E., Vitiello, G. (2004). Quantum noise induced entanglement and chaos in the dissipative quantum model of brain, *Int. J. Mod. Phys. B*, **18** 841–858.
- Petersen, P. (1998). *Riemannian Geometry*. Springer, New York.
- Petersen, P. (1999). Aspects of Global Riemannian Geometry. *Bull. Amer. Math. Soc.*, **36**(3), 297–344.
- Peterson, I. (1993). *Newton’s Clock: Chaos in the Solar System*. W. H. Freeman, San Francisco.
- Peyrard, M., Bishop, A.R. (1989). Statistical mechanics of a nonlinear model for DNA denaturation, *Phys. Rev. Lett.* **62**, 2755.
- Piaget, J., Henriques, G., Ascher, E. (1992). *Morphisms and categories*. Erlbaum Associates, Hillsdale, NJ.
- Pitt, M.A., Myung, I.J., Zhang, S. (2002). Toward a method of selecting among computational models of cognition *Psychological Review*, **109**(3), 472–491.
- PlanetMath.org (2006).
<http://planetmath.org/encyclopedia/TopologicalSpace.html>
- Poincaré, H. (1890). Sur le problème differential equations trois corps et les équations de la dynamique. *Acta Math.*, **13**, 1–270.
- Polchinski, J. (1995). Dirichlet-branes and Ramond-Ramond charges, *Phys. Rev.*

- Lett. **75**, 4724.
- Polchinski, J. (1998). *String theory*. 2 Vols. Cambridge Univ. Press, Cambridge, UK.
- Pommaret, J. (1978). *Systems of Partial Differential Equations and Lie Pseudogroups*, Gordon and Breach, Glasgow.
- Ponomarev, V., Obukhov, Yu. (1982). Generalized Einstein-Maxwell theory. *GRG* **14**, 309.
- Pons, J.M., Salisbury, D.C., Shepley, L.C. (2000). Gauge Transformations in Einstein-YM Theories. *J.Math.Phys.* **41**, 5557-5571.
- Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., Mishchenko, E.F. (1986). *The Mathematical Theory of Optimal Processes*, Gordon & Breach Science Publishers, New York.
- Ponzano, G., Regge, T. (1968). Spectroscopy and Group Theoretical Methods. in F. Block (ed) *Physics*, North Holland, New York.
- Popper, K.R. (1959). *The Logic of Scientific Discovery*. New York, NY: Basic Books.
- Postnikov, M.M. (1986). *Lectures in Geometry V, Lie Groups and Lie Algebras*. Mir Publ., Moscow.
- Potters, M., Bouchaud, J.P., Sestovic, D. (2001). Hedged Monte-Carlo: low variance derivative pricing with objective probabilities. *Physica A* **289**, 517-25.
- Perelomov, A.M. (1990). *Integrable systems of classical mechanics and Lie algebras*, I. Birkhäuser.
- Principe, J., Euliano, N., Lefebvre, C. (2000). *Neural and Adaptive Systems: Fundamentals Through Simulations*. Wiley, New York.
- Prykarpatsky, A.K. (1996). Geometric models of the Blackmore's swept volume dynamical systems and their integrability, In: *Proc. of the IMACS-95, Hamburg 1995*, ZAMP, **247**(5), 720-724.
- Pulvermüller, F. (2005). Brain mechanisms kinking language and action. *Nature Rev. Neurosci.* **6**, 576-582.
- Putra, M. (1993). *Hamiltonian Mechanical Systems and Geometric Quantization*, Kluwer, Dordrecht.
- Ramond, P. (1990). *Field Theory: a Modern Primer*. Addison-Wesley, Reading.
- Rani, R., Edgar, S.B., Barnes, A. (2003). Killing Tensors and Conformal Killing Tensors from Conformal Killing Vectors. *Class.Quant.Grav.*, **20**, 1929-1942.
- Rashevsky, N. (1967). Organismic sets. Outline of a general theory of biological and sociological organisms. *Bull. Math. Biophys.*, **29**, 139-152.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psych. Rev.*, **85**, 59-108.
- Reed, M., Simon, B. (1975). *Methods of modern mathematical physics, Vol. 2: Fourier analysis, self-adjointness*. Academic Press, San Diego.
- Regge, T. (1961). General relativity without coordinates. *Nuovo Cim. A* **19**, 558-571.
- Reidemeister, K. (1948). *Knot Theory*, Chelsea Pub., New York.
- Reisenberger, M., Rovelli, C. (1997). 'Sum over surfaces' form of loop quantum gravity. *Phys. Rev. D* **56**, 3490-3508.
- Reshetikhin, N., Turaev, V. (1990). Ribbon graphs and their invariants derived from quantum groups. *Comm. Math. Phys.* **127**, 1-26.

- Reshetikhin, N.Yu, Turaev, V.G. (1991). Invariants of four-manifolds via link polynomials and quantum groups. *Invent. Math.* **10**, 547.
- Reznikov, A.G. (1997). Characteristic classes in symplectic topology, *Selecta Math.* **3**, 601–642.
- De Rham, G. (1984). *Differentiable Manifolds*. Springer, Berlin.
- Ricca, R.L. (1993). Torus knots and polynomial invariants for a class of soliton equations, *Chaos*, **3**(1), 83–91.
- Ricciardi, L.M., Umezawa, H. (1967). Brain physics and many-body problems, *Kibernetik*, **4**, 44.
- Ritter, H., Martinetz, T., Schulten, K. (1992). *Neural Computation and Self-Organizing Maps: An Introduction*, Addison–Wesley, Reading.
- Roe, J. (1988). *Elliptic operators, Topology and Asymptotic Methods*, Longman Sc.
- Roe, R.M., Busemeyer, J.R., Townsend, J.T. (2001). Multi-alternative decision field theory: A dynamic connectionist model of decision making. *Psych. Rev.*, **108**, 370–392.
- Rosa-Clot, M., Taddei, S. (2002). A Path Integral Approach to Derivative Security Pricing. *Int. J. Theor. Appl. Finance*, **5**(2), 123–146.
- Rosquist, K. (1980). Global rotation. *Gen. Rel. Grav.*, **12**, 649.
- Rosquist, K. (1989). Killing tensor conservation laws and their generators. *J. Math. Phys.*, **30**, 2319–2321.
- Rosquist, K., Pucacco, G.. (1995). Invariants at fixed and arbitrary energy. A unified geometric approach. *J. Phys. A: Math. Gen.*, **28**, 3235–3252.
- K. Rosquist. (1997). A tensorial lax pair equation and integrable systems in relativity and classical mechanics. In R.T. Jantzen and G.M. Keiser, editors, *On recent developments in Theoretical and Experimental General Relativity, Gravitation and Relativistic Field Theories*, Vol. I, 379. World Scientific, Singapore.
- Rosquist, K., Goliath, M. (1997). Lax pair tensors and integrable space-times. [arXiv:gr-qc/9707003](https://arxiv.org/abs/gr-qc/9707003).
- Rovelli, C. (1993). The basis of the Ponzano-Regge-Turaev-Viro-Ooguri model is the loop representation basis. *Phys. Rev.* **D48**, 2702–2707.
- Rovelli, C. (1996). Black Hole Entropy from Loop Quantum Gravity, *Phys. Rev. Lett.* **14**, 3288–3291.
- Rovelli, C. (1996). Loop Quantum Gravity and Black Hole Physics, *Helv. Phys. Acta.* **69**, 582–611.
- Rovelli, C. (1997). Strings, loops and others: a critical survey of the present approaches to quantum gravity. Plenary lecture on quantum gravity at the GR15 conference, Puna, India.
- Rovelli, C., Smolin, L. (1988). Knot theory and quantum gravity, *Phys. Rev. Lett.* **61**, 1155.
- Rovelli, C., Smolin, L. (1994). The physical Hamiltonian in non-perturbative quantum gravity, *Phys. Rev. Lett.* **72**, 446.
- Rovelli, C. (1997). Half way through the woods, in Earman, J, and Norton, J, (eds.) *The Cosmos of Science*, 180–223, Univ. Pittsburgh Press, Konstanz.
- Rovelli, C., Smolin, L. (1987). A new approach to quantum gravity based on loop

- variables. Int. Conf. Gravitation and Cosmology, Goa, Dec 14-19 India.
- Rovelli, C., Smolin, L. (1990). Loop representation of quantum general relativity, Nucl. Phys., **B331**:(1), 80–152.
- Rovelli, C., Smolin, L. (1995). Spin Networks and Quantum Gravity, Phys. Rev. **D52**, 5743–5759.
- Rovelli, C. (1998). Loop quantum gravity. Living Rev. Rel. **1**.
- Ruelle, D. (1978). Thermodynamic formalism. Encyclopaedia of Mathematics and its Applications. Addison–Wesley, Reading.
- Rumelhart, D., McClelland, J. (eds.) (1987). Parallel Distributed Processing. Vols I & II, MIT Press, Cambridge.
- Russell, N. (2003). Artificial Intelligence. A modern Approach. Prentice Hall.
- Ryder, L. (1996). Quantum Field Theory. Cambridge Univ. Press, Cambridge, UK.
- Sardanashvily, G., Zakharov, O. (1992). Gauge Gravitation Theory. World Scientific, Singapore.
- Sardanashvily, G. (1992). On the geometry of spontaneous symmetry breaking. J. Math. Phys. **33**, 1546.
- Sardanashvily, G. (1993). Gauge Theory in Jet Manifolds. Hadronic Press, Palm Harbor, FL.
- Sardanashvily, G. (1994). Energy-momentum conservation law in Hamiltonian field theory, arXiv:gr-qc/9412041.
- Sardanashvily, G. (1994). Constraint field systems in multimomentum canonical variables. J. Math. Phys. **35**, 6584.
- Sardanashvily, G. (1995). Generalized Hamiltonian Formalism for Field Theory. Constraint Systems., World Scientific, Singapore.
- Sardanashvily, G. (1997). Stress-energy-momentum tensors in constraint field theories. J. Math. Phys. **38**, 847.
- Sardanashvily, G. (1998). Hamiltonian time-dependent mechanics. J. Math. Phys. **39**, 2714.
- Sardanashvily, G. (1998). Stress-Energy-Momentum Tensors in Constraint Field Theories. J. Math. Phys. **39**, 2714.
- Sardanashvily, G. (2002). Ten lectures on jet manifolds in classical and quantum field theory. arXiv:math-ph/0203040.
- Sardanashvily, G. (2002). The Lyapunov stability of first order dynamic equations with respect to time-dependent Riemannian metrics. arXiv nlin.CD/0201060.
- Sardanashvily, G. (2002). The bracket and the evolution operator in covariant Hamiltonian field theory, arXiv:math-ph/0209001.
- Sardanashvily, G. (2003). Geometric quantization of relativistic Hamiltonian mechanics. Int. J. Theor. Phys. **42**, 697-704.
- Sastri, S.S., Isidori, A. (1989). Adaptive control of linearizable systems. IEEE Trans. Aut. Con., **34**(11), 1123–1131.
- Satarić, M.V., Tuszynski, J.A. and Zakula, R.B. (1993). Kinklike excitations as an energy-transfer mechanism in microtubules. Phys. Rev. E, **48**, 589–597.
- Satarić, M.V., Zeković, S., Tuszynski, J.A., Pokorný, J. (1998). Mössbauer effect as a possible tool in detecting nonlinear excitations in microtubules. Phys.

- Rev. E, **58**, 6333–6339.
- Saunders, D.J. (1989). *The Geometry of Jet Bundles*. Lond. Math. Soc. Lect. Notes Ser. **142**, Cambr. Univ. Pr.
- Schaal, S. (1998). Robot learning. In M. Arbib (ed). *Handbook of Brain Theory and Neural Networks* (2nd ed.), MIT Press, Cambridge.
- Schaal, S., Atkeson, C.G. (1998). Constructive incremental learning from only local information. *Neural Comput.*, **10**, 2047–2084.
- Schaal, S. (1999). Is imitation learning the route to humanoid robots?. *Trends Cogn. Sci.*, **3**, 233–242.
- Schaal, S., Atkeson, C.G. (1998). Constructive incremental learning from only local information. *Neural Comput.*, **10**, 2047–2084.
- Schaal, S., Sternad, J.C., Osu, R., Kawato, M. (2004). Rhythmic arm movement is not discrete. *Nature Neurosci.* **7**, 1136–1143.
- Schafer, R.D. (1996). *An Introduction to Nonassociative Algebras*. Dover, New York.
- Schaub, H., Tsiotras, P., Junkins, J.: Principal Rotation Representations of Proper $N \times N$ Orthogonal Matrices. *Int. J. Eng. Sci.*, **33**, (15), 2277–2295, (1995)
- Schellekens, A.N.: Introduction to conformal field theory. *Fortsch. Phys.*, **44**, 605, (1996)
- Schiff, S.J., Jerger, K., Duong, D.H., Chang, T., Spano, M.L., Ditto, W.L. (1994). Controlling chaos in the brain, *Nature*, **370**, 615–620.
- Schulman, L.S. (1981). *Techniques and Applications of Path Integration*. New York, Wiley.
- Schwarz, A.S. (1978). New topological invariants in the theory of quantised fields. *Lett. Math. Phys.* **2**, 247.
- Schwarz, M. (1993). *Morse Homology*. Birkhäuser, Basel.
- Seiberg, N., Witten, E. (1994). Monopoles, duality and chiral symmetry breaking in $N = 2$ supersymmetric QCD. *Nucl. Phys.* **B426**, 19–52.
- Seiberg, N., Witten, E. (1994). Electric-magnetic duality, monopole condensation, and confinement in $N = 2$ supersymmetric Yang–Mills theory. *Nucl. Phys.* **431**, 484–550.
- Seiberg, N., Witten, E. (1999). String Theory and Noncommutative Geometry, *JHEP* **9909**, 032.
- Seidel, P. (1997). π_1 of symplectic automorphism groups and invertibles in quantum cohomology rings, *Geometric and Func. Anal.* **7**, 1046–95.
- Seiler, W.M. (1995). Involution and constrained dynamics II: The Faddeev–Jackiw approach, *J. Phys. A*, **28**, 7315–7331.
- Sejnowski, T., Koch, C., Churchland, P. (1988). Computational Neuroscience. *Science*, **241**, 1299–1306.
- Semmelmann, U. (2002). Conformal Killing forms on Riemannian manifolds. *arXiv math.DG/0206117*.
- Sen, A. (1998). Tachyon Condensation On The Brane–Antibrane System. *JHEP* **98**, 8–12.
- Seraji, H. (1989). Configuration control of redundant manipulators: theory and implementation. *IEEE Trans. Rob. Aut.*, **5**, 437–443.

- Shabanov, S.V., Klauder, J.R. (1998). Path Integral Quantization and Riemannian-Symplectic Manifolds. *Phys. Lett. B* **435**, 343–349.
- Shen, Z. (2005). Riemann–Finsler geometry with applications to information geometry, unpublished material.
- Shih, C.L., Gruver, W., Lee, T. (1993). Inverse kinematics and inverse dynamics for control of a biped walking machine. *J. Robot. Syst.*, **10**, 531–555.
- Shih, C.L., Klein, C.A. (1993). An adaptive gait for legged walking machines over rough terrain. *IEEE Trans. Syst. Man, Cyber. A*, **23**, 1150–1154.
- Shuster, M.D. (1993). A Survey of Attitude Representations, *J. Astr. Sci.*, **41**(4), 316–321.
- Siegel, W. (2002). *Fields* (2nd Ed.), <http://insti.physics.sunysb.edu/siegel/plan.html>,
- Skiffington, S., Fernandez, E., McFarland, K. (1998). Towards a validation of multiple features in the assessment of emotions. *Eur. J. Psych. Asses.* **14**(3).
- van der Smagt, P. (ed.) (1999). *Self-Learning Robots. Workshop: Brainstyle Robotics*, IEE, London.
- Smale, S. (1960). The generalized Poincaré conjecture in higher dimensions, *Bull. Amer. Math. Soc.*, **66**, 373–375.
- Smale, S. (1967). Differentiable dynamical systems, *Bull. Amer. Math. Soc.*, **73**, 747–817.
- Smolin, L. (1997). *Loops and Strings*, Living Reviews, September, 1997.
- Smolin, L. (1997). *The Life of the Cosmos*. Oxford Univ. Press, Oxford.
- Sniatycki, J. (1980). *Geometric Quantization and Quantum Mechanics*. Springer-Verlag, Berlin.
- Snyder, H.S. (1947). Quantized Space-Time. *Phys. Rev.* **71**, 38.
- Snyder, H.S. (1947). The Electromagnetic Field in Quantized Space-Time. *Phys. Rev.* **72**, 68.
- Sober, S.J., Sabes, P.N. (2003). Multisensory integration during motor planning. *J. Neurosci.* **23**(18), 6982–92.
- Sober, S.J., Sabes, P.N. (2005). Flexible strategies for sensory integration during motor planning. *Nature Neurosci.* **8**, 490–497.
- Socolovsky, M. (1991). Gauge transformations in fibre bundle theory. *J. Math. Phys.*, **32**, 2522.
- Sparrow, C. (1982). *The Lorenz Equations: Bifurcations, Chaos and Strange Attractors*. Springer, New York.
- Sperry, R.W. (1969). A modified concept of consciousness. *Psych. Rev.*, **76**, 532–536.
- Spivak, M. (1965). *Calculus on Manifolds: A Modern Approach to Classical Theorems of Advanced Calculus*. HarperCollins Publishers.
- Spivak, M. (1970–75). *A comprehensive introduction to differential geometry*, Vol. I–V, Publish or Perish Inc., Berkeley.
- Srivastava, Y.N., Vitiello, G., Widom, A. (1995). Quantum dissipation and quantum noise, *Annals Phys.*, **238**, 200.
- Srivastava, Y.N., Widom, A., Vitiello, G. (1999). Quantum measurements, information and entropy production. *Int. J. Mod. Phys. B13*, 3369–3382.
- Stapp, H.P., (1993). *Mind, Matter and Quantum Mechanics*. Springer, New York.

- Stasheff, J.D. (1963). Homotopy associativity of H-spaces I & II. *Trans. Amer. Math. Soc.*, **108**, 275–292, 293–312.
- Steenrod, N. (1951). *The Topology of Fibre Bundles*. Princeton Univ. Press, Princeton.
- Stengel, R. (1993). *Optimal control and estimation*. Dover, New York.
- Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *Am. Sci.*, **57**(4), 421–457.
- Stevens, B.L., Lewis, F.L. (2003). *Aircraft Control and Simulation* (2nd ed.). Wiley, Hoboken, NJ.
- Stong, R.E. (1968). *Notes on Cobordism Theory*. Princeton Univ. Press, Princeton.
- Strogatz, S. (2000). From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D*, **143**, 1–20.
- Strominger, A. (1990). Special Geometry. *Commun. Math. Phys.* **133**, 163.
- Strominger, A., Vafa, C. (1996). Microscopic Origin of the Bekenstein-Hawking Entropy, *Phys. Lett.* **B379**, 99–104.
- Stuart, J. (1999). *Calculus* (4th ed.). Brooks/Cole Publ., Pacific Grove, CA.
- Stuart, C.I.J., Takahashi, Y., Umezawa, H. (1978). On the stability and non-local properties of memory, *J. Theor. Biol.* **71**, 605–618.
- Stuart, C.I.J., Takahashi, Y., Umezawa, H. (1979). Mixed system brain dynamics: neural memory as a macroscopic ordered state, *Found. Phys.* **9**, 301.
- Sugino, F. (2000). Witten's open string field theory in constant B-field background. *J. High Energy Phys.* JHEP03(2000)017.
- Sulanke, R., Wintgen, P. (1972). *Differential geometry und faser-bundel*; bound 75, Veb. Deutscher Verlag der Wissenschaften, Berlin.
- Sussmann, H.J. (1983). Lie brackets and local controllability: a sufficient condition for scalar-input systems, *SIAM J. Con. Opt.*, **21**(5), 686–713.
- Sussmann, H.J. (1987). A general Theorem on local controllability, *SIAM J. Con. Opt.*, **25**(1), 158–194.
- Sutton, R.S., Barto, A.G. (1998). *Reinforcement Learning: An Introduction*. MIT Press, Cambridge.
- Switzer, R.K. (1975). *Algebraic Topology – Homology and Homotopy*. (in *Classics in Mathematics*), Springer, New York.
- Tabony, J., Vuillard, L., Papaseit, C. (1999). Biological self-organisation and pattern formation by way of microtubule reaction-diffusion processes. *Adv. Complex Syst.* **2**(3), 221–276.
- Tabuada, P., Pappas, G.J. (2001). *Abstractions of Hamiltonian Control Systems*. Proceedings of the 40th IEEE Conf. Decis. Con., Orlando, FL.
- Taiji, M., Ikegami, T. (1999). Dynamics of internal models in game players. *Physica D*, **134**, 253–266.
- Tanaka, K., Wang, H.O. (2001). *Fuzzy control systems design and analysis*. Wiley, New York.
- Taubes, C.H. (1990). Casson's invariant and gauge theory, *J. Diff. Geom.* **31**, 547–599.
- Taubes, C.H. (1994). The Seiberg-Witten invariants and symplectic forms, *Math. Res. Lett.* **1**, 809–822.

- Taubes, C.H. (1995). More constraints on symplectic manifolds from Seiberg-Witten invariants, *Math. Research Letters*, **2**, 9-14;
- Taubes, C.H. (1995). The Seiberg-Witten invariants and The Gromov invariants, *Math. Research Letters* **2**, 221-238;
- Teitelboim, C. (1983). Causality versus gauge invariance in quantum gravity and supergravity. *Phys. Rev. Lett.* **50**, 705-708.
- T. Thiemann, Anomaly-Free Formulation of Nonperturbative Four-Dimensional Lorentzian Quantum Gravity, *Phys. Lett.* **B380** (1996) 257-264.
- Thirring, W. (1979). *A Course in Mathematical Physics* (in four volumes). Springer, New York.
- Thom, R. (1975). *Structural Stability and Morphogenesis*. Addison-Wesley, Reading.
- Thompson, G. (1993). Non-uniqueness of metrics compatible with a symmetric connection. *Class. Quant. Grav.* **10**, 2035.
- Thorpe, J.A. (1979). *Elementary Topics in Differential Geometry*. Springer, New York.
- Toda, M. (1981). *Theory of Nonlinear Lattices*. Springer, New York.
- Toda, M., Kubo, R., Saito, N. (1991). *Statistical Physics 2*, Springer, New York.
- Tomlin, C., Pappas, G.J., Sastry, S. (1998). Conflict resolution for air traffic management: A case study in multi-agent hybrid systems, *IEEE Trans. Aut. Con.*, **43**, 509-521.
- Townsend, J.S. (1992). *A modern Approach to Quantum Mechanics*, McGraw Hill Inc., New York.
- Tsue, Y., Fujiwara, Y. (1991). Time-Dependent Variational Approach to (1+1)-Dimensional Scalar-Field Solitons, *Progr. Theor. Phys.* **86**(2), 469-489.
- Tucker, R., Wang, C. (1995). Black holes with Weyl charge and non-Riemannian waves. *Class. Quant. Grav.* **12**, 2587.
- Turaev, V., Viro, O. (1992). State Sum Invariants of 3- Manifolds and Quantum 6-J Symbols. *Topology* **31**, 865-902.
- Udriste, C. (2000). *Geometric dynamics*. Kluwer Academic Publishers.
- Udriste, C., Neagu, M. (1999). Extrema of p -energy functional on a Finsler manifold. *Diff. Geom. Dyn. Sys.* **1**(1), 10-19.
- Vafa, C., Witten, E. (1994). A strong coupling test of S duality, arXiv:hep-th/9408074.
- Vaisman, I. (1994). *Lectures on the Geometry of Poisson Manifolds*. Birkhäuser Verlag, Basel.
- Veneziano, G. (1991). Scale factor duality for classical and quantum strings. *Phys. Lett.* **B265**, 287.
- Verlinde, E. (1988). Fusion rules and modular transformations in 2D conformal field theory. *Nucl. Phys.* **B300**, 360-376.
- Vitiello, G. (1995). Dissipation and memory capacity in the quantum brain model, *Int. J. Mod. Phys. B*, **9**, 973-989.
- Vitiello, G. (2001). *My Double Unveiled*, John Benjamins, Amsterdam.
- Vladimirov, V.S., (1971). *Equations of Mathematical Physics*. Marcel Dekker, Inc., New York.
- Vladimirov, V.S. (1986). *A Collection of Problems on the Equations of Mathe-*

- mathematical Physics. Springer, Berlin.
- Voisin, C. (2002). *Hodge Theory and Complex Algebraic Geometry I*. Cambridge Univ. Press, Cambridge.
- Vonk, M. (2005). A mini-course on topological strings. arXiv:hep-th/0504147.
- Vukobratovic, M. (1970). On the stability of biped locomotion. *IEEE Trans. Biom. Eng.*, **17**, 25–36.
- Vukobratovic, M., Juricic, D., Frank, A. (1970). On the control and stability of one class of biped locomotion systems. *ASME J. Basic Eng.*, **92**, 328–332.
- Vukobratovic, M., Stepanenko, Y. (1972). On the stability of anthropomorphic systems. *Math. Biosci.* **15**, 1–37.
- Vukobratovic, M., Stepanenko, Y. (1973). Mathematical models of general anthropomorphic systems. *Math. Biosci.* **17**, 191–242.
- Vukobratovic, M. (1975). *Legged Locomotion Robots and Anthropomorphic Mechanisms*. Mihailo Pupin Institute, Belgrade.
- Vukobratovic, M., Stokic, D.: (1982). *Scientific Fundamentals of Robotics, Vol. 2, Control of Manipulation Robots: Theory and Application*, Springer-Verlag, Berlin.
- Vukobratovic, M., Borovac, B., Surla, D., Stokic, D., (1990). *Biped Locomotion: Dynamics, Stability, Control, and Applications*. Springer, Berlin.
- Vygotsky, L.S. (1982). Historical meaning of the Psychological crisis. *Collected works. vol. 1*. Pedag. Publ., Moscow.
- Wald, R. (1994). *Quantum field theory in curved space-time and black hole thermodynamics*, Chicago Univ. Press.
- Walker, M., Penrose, R. (1970). On quadratic first integrals of the geodesic equations for type $\{2,2\}$ space-times. *Comm. Math. Phys.*, **18**, 265–274.
- Wall, M.M., Amemiya, Y. (1998). Fitting nonlinear structural equation models. *Proc. Social Stat. Section. Ann. Meet. Ame. Stat. Assoc.* 180–185.
- Wall, M.M., Amemiya, Y. (2000). Estimation for polynomial structural equation models. *Journal of American Statistical Association*, **95**, 929–940.
- Watson, L.T. (1990). Globally convergent homotopy algorithms for nonlinear systems of equations. *Nonlinear Dynamics*, **1**, 143–191.
- Watts, D.J., Strogatz, S.H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, **393**, 440–442.
- Wehner, M.F., Wolfer, W.G. (1983). Numerical evaluation of path-integral solutions to Fokker–Planck equations. I., *Phys. Rev. A* **27**, 2663–2670.
- Wehner, M.F., Wolfer, W.G. (1983). Numerical evaluation of path-integral solutions to Fokker–Planck equations. II. Restricted stochastic processes, *Phys. Rev. A*, **28**, 3003–3011.
- Weinberg, S. (1964). Photons and Gravitons in S-Matrix Theory: Derivation of Charge Conservation and Equality of Gravitational and Inertial Mass. *Phys. Rev.* **B135**, 1049.
- Weinberg, S. (1980). Conceptual foundations of the unified theory of weak and electromagnetic interactions. *Rev. Mod. Phys.* **52**, 515–523.
- Weinstein, A. (1990). Affine Poisson structures. *Internat. J. Math.*, **1**, 343–360.
- Weisstein, E.W. (2004). *MathWorld—A Wolfram Research Web Resource*. <http://mathworld.wolfram.com>.

- Wess, J., Zumino, B. (1991). Covariant Differential Calculus on the Quantum Hyperplane, *Nucl. Phys. Proc. Suppl.* **18B**, 302.
- Wess, J. (2004). Deformed coordinate spaces: Derivatives, arXiv:hep-th/0408080.
- West, P. (1990). *Introduction to Supersymmetry and Supergravity*. World Scientific, Singapore.
- Weyl, H. (1927). *Quantenmechanik und Gruppentheorie*. *Zeitschrift für Physik*, **46**, 1-46.
- Wheeler, J.A. (1961). Geometrodynamics and the Problem of Motion. *Rev. Mod. Phys.* **33**, 63–78.
- Wheeler, J.A. (1962). *Geometrodynamics*. Academic Press, New York.
- Wheeler, J.A., Feynman, R.P. (1949). Classical Electrodynamics in Terms of Direct Interparticle Action, *Rev. Mod. Phys.*, **21**, 425–433.
- Whitney, D.E. (1987). Historical perspective and state of the art in robot force control. *Int. J. Robot. Res.*, **6**(1), 3–14.
- Wiener, N. (1961). *Cybernetics*. Wiley, New York,
- Wiggins, S. (1990). *Introduction to Applied Dynamical Systems and Chaos*. Springer, New York.
- Wikipedia, the free encyclopedia, (2005) <http://wikipedia.org>.
- Williams, R.M., Tuckey, P. (1992). Regge Calculus: A bibliography and brief review. *Class. Quant. Grav.* **9**, 1409.
- Williams, R.M. (1997). Recent progress in Regge calculus. *Nucl. Phys. B* **57**, 73–81.
- Williams, C.P., Clearwater, S.H. (1998). *Explorations in Quantum Computing*, Springer, New York.
- Willmore, T.J. (1993). *Riemannian Geometry*. Oxford Univ. Press, Oxford.
- Wilmott, P., Dewynne, J., Howinson, S. (1993). *Option Pricing: Mathematical Models and Computation*. Oxford Financial Press.
- Wilson, D. (2000). *Nonlinear Control*, Advanced Control Course (Student Version), Karlstad Univ.
- Winter, D.A. (1990). *Biomechanics and Motor Control of Human Movement*, (2nd ed.) Wiley, New York.
- Witten, E. (1982). Supersymmetry and Morse theory. *J. Diff. Geom.*, **17**, 661–692.
- Witten, E. (1986). Interacting Field Theory of Open Superstrings, *Nucl. Phys.* **B276**, 291.
- Witten, E. (1986). Non-Commutative Geometry and String Field Theory, *Nucl. Phys.* **B268**, 253.
- Witten, E. (1988). Topological Quantum Field Theory, *Commun. Math. Phys.* **117**, 353-386;
- Witten, E. (1988). Topological Sigma Models. *Commun. Math. Phys.* **118**, 411.
- Witten, E. (1988). 2+1 Gravity as an Exactly Soluble Model. *Nucl. Phys.* **B311**, 46-78.
- Witten, E. (1988). Topological Gravity. *Phys. Lett. B* **206**, 601.
- Witten, E. (1989). Quantum field theory and the Jones polynomial. *Commun. Math. Phys.* **121**, 351.
- Witten, E. (1990). On the structure of the topological phase of two-dimensional gravity. *Nucl. Phys. B* **340**, 281.

- Witten, E. (1991). Introduction to Cohomological Field Theories. *Int. J. Mod. Phys. A* **6**, 2775.
- Witten, E. (1992). Mirror manifolds and topological field theory. In *Essays on mirror manifolds*, ed. S.-T. Yau, International Press, 120–158.
- Witten, E. (1994). Monopoles and four manifolds. *Math. Res. Lett.* **1**, 769–796.
- Witten, E. (1995). Chern–Simons gauge theory as a string theory. *Prog. Math.* **133**, 637.
- Witten, E. (1995). String Theory Dynamics in Various Dimensions. *Nucl. Phys. B* **443**, 85.
- Witten, E. (1996). Bound States of Strings and p-Branes. *Nucl. Phys.* **B460**, 335.
- Witten, E. (1998). Magic, mystery, and matrix. *Notices AMS*, **45**(9), 1124–1129.
- Witten, E. (1998). Anti-de Sitter space and holography. *Adv. Theor. Math. Phys.*, **2**, 253.
- Witten, E. (1998). D –Branes and K –Theory. *JHEP* **98**, 12–19.
- Witten, E. (2000). Overview of K-Theory Applied To Strings. *arXiv:hep-th/0007175*.
- Witten, E. (2002). The Universe on a String. *Astronomy magazine*, June, 2002.
- Woodhouse, N. (1992). *Geometric Quantization*. Clarendon Press, Oxford.
- Xu, Z., Hauser, J. (1994). Higher order approximate feedback linearization about a manifold, *J. Math. Sys. Est. Con.*, **4**, 451–465.
- Xu, Z., Hauser, J. (1995). Higher order approximate feedback linearization about a manifold for multi-input systems, *IEEE Trans. Aut. Con*, **AC-40**, 833–840.
- Yager, R.R. (1987). *Fuzzy Sets and Applications: Selected Papers by L.A. Zadeh*, Wiley, New York.
- Yalcin, I., Amemiya, Y. (2001). Nonlinear factor analysis as a statistical method. *Statistical Science*, **16**, 275–294.
- Yang, C.N., Lee, T.D. (1952). Statistical theory of equation of state and phase transitions I: Theory of condensation. *Phys. Rev.* **87**, 404–409.
- Yang, J.V. (1995). Introduction to Seiberg-Witten’s Invariants. *arXiv:dg-ga/9508005*.
- Yano, K. (1952). On harmonic and Killing vector fields. *Ann. Math.* **55**(2), 328–347.
- Yau, S.-T. (1978). On the Ricci curvature of a compact Kähler manifold and the complex Monge-Ampère equation, I. *Comm. Pure Appl. Math.* **31**, 339–411.
- Yong, J., Zhou, X. (1999). *Stochastic controls. Hamiltonian Systems and HJB Equations*. Springer, New York.
- Yorke, J.A., Alligood, K., Sauer, T. (1996). *Chaos: An Introduction to Dynamical Systems*. Springer, New York.
- Zakharov, O. (1992). Hamiltonian formalism for nonregular Lagrangian theories in fibred manifolds. *J. Math. Phys.* **33**, 607.
- Zhang, R.B., Gould, M.D., Bracken, A.J. (1991). Quantum Group Invariants and Link Polynomials. *Commun. Math. Phys.* **137**, 13.
- Zhang, R.B., Wang, B.L., Carey, A.L., McCarthy, J. (1995). Topological Quantum Field Theory and Seiberg-Witten Monopoles. *arXiv:hep-th/9504005*.

This page intentionally left blank

Index

- 1-jet bundle, 827
- 1-jet lift, 805
- 1-jet space, 800, 802
- 2-jet lift, 820
- 2-jet space, 800, 819
- c -split, 650
- k -jet, 797
- k -jet space, 822

- absolute covariant derivative, 273, 280
- absolute time derivative, 156
- action, 23, 24, 83, 1098, 1181, 1194
- action of a Lie group, 207
- action paradigm, 87
- action principle, 24, 84
- action-amplitude picture, 983, 1079
- action-angle variables, 344, 365, 370
- adaptive path integral, 997, 1080, 1082, 1083
- adjoint action, 671
- adjoint bundle, 917
- adjoint functors, 111
- adjoint Jacobi equation, 572
- admissible Dynkin diagrams, 242
- affine 1-jet bundle, 827
- affine bundle, 500
- affine connection, 19, 292, 808
- affine control system, 556
- affine information structure, 334
- affine jet bundle, 803
- AG-structures, 693
- agent's momentum phase-space, 603
- agent's velocity phase-space, 603
- algebra homomorphism, 202
- algebra of noncommutative functions, 460
- algebraic cycle theory, 511
- algebraic K -theory, 510
- algebraic properties, 28
- all, 638
- almost complex structure, 656
- Ambrose-Singer Theorem, 269
- American option, 1035
- Amid I peptide groups, 381
- amplitude, 970, 991
- analytic map, 28
- anti-self-dual basis, 681
- antipode, 463
- any, 649, 650
- approximate feedback linearization, 545
- arc-element, 275
- area functional, 877
- Arnold conjecture, 28
- Ashtekar connection, 709
- Ashtekar phase-space formalism, 708
- associated graded algebra, 671
- associativity of morphisms, 103
- associator, 133
- Atiyah axioms, 302
- Atiyah-Bott theorem, 653
- Atiyah-Singer Index Theorem, 518
- atlas, 2, 6, 144
- attracting set, 342

- automorphism of the Hamiltonian structure, 629
- autonomous agents, 603
- Axioms of Lie super-algebra, 671
- background fields, 887
- Baker–Campbell–Hausdorff formula, 29, 461
- Banach manifold, 147
- Banach space, 146
- basin of attraction, 341
- Batalin–Vilkovisky quantization, 1119, 1125
- Bekenstein–Hawking entropy, 1227
- Bekenstein–Hawking formula, 294, 709
- Bendixon criterion, 343
- Betti numbers, 5, 28, 105, 188, 1076
- Bianchi covariant derivative, 290, 379
- Bianchi identity, 21, 35, 528, 743, 913
- Bianchi relation, 124
 - dual, 124
 - extended, 125
- Bianchi symmetry condition, 278
- biholomorphism, 429
- bijection, 2
- black hole dynamics, 1225
- Black–Scholes–Merton formula, 1025
- blow-down map, 653
- body-fixed frame, 354
- Bogomol’nyi monopoles, 1119
- Bolza problem, 570
- Borel–Hirzebruch theorem, 653
- Bose–Einstein condensate, 992
- Bose–Einstein statistics, 34
- bosonic string theory, 40, 1146
- Bott Periodicity, 494
- Bott Periodicity Theorem, 509, 516
- Bott tower, 653
- boundary operator, 183
- BPS formula, 736
- braided monoidal category, 129
- brain-like control functor, 573
- branched polymer, 301
- brane, 1143, 1182
- brane world-sheet, 1171
- brane-world scenario, 37
- brane-world theory, 45
- Brouwer degree, 284
- Brownian dynamics, 986
- Brownian motion, 1026
- BRST quantization, 1126
- BRST symmetry, 1054, 1197
- BRST-operator, 1105, 1120
- bulk, 37, 41
- bundle diffeomorphism, 490
- Burgers dynamical system, 268
- Calabi–Yau manifold, 45, 436
- Calabi–Yau manifolds, 1180, 1186, 1187
- calculus of variations, 23, 84
- calibrated submanifolds, 435
- calibration form, 435
- Campbell–Baker–Hausdorff formula, 195
- canonical 3-form, 843
- canonical lift, 608
- canonical momenta indicators, 1019
- canonical Poisson structure, 843
- canonical polysymplectic form, 905
- canonical soldering form, 612
- canonical tangent-valued 1-form, 611
- canonical transformation, 336
- Cartan equations, 842
- Cartan magic formula, 198
- Cartan relation, 610
- Cartan subalgebra, 242, 734
- Cartan Theorem, 263
- Cartesian product, 4
- Casimir effect, 45
- Casimir form, 368
- Casimir function, 775
- Casson invariant, 1119
- categorification, 127
- category, 28, 116
- category of Hopf algebras, 462
- Cauchy Theorem, 351, 1208
- Cauchy–Riemann equation, 655
- Cauchy–Riemann equations, 428
- cellular chain complex, 640
- chain rule, 182

- Chan–Paton factor, 1238
- Chan–Paton gauge bundle, 1242
- change of coordinates, 7
- chaotic behavior, 506
- Chapman–Kolmogorov equation, 988, 1030
- Chapman–Kolmogorov
 - integro–differential equation, 989
- Chapman–Kolmogorov law, 165, 173
- characteristic classes, 511
- chart, 2, 6
- Cech cohomology, 509
- Chern character, 511, 517, 522
- Chern class, 761, 1122, 1133, 1192, 1230
- Chern classes, 511, 516
- Chern form, 512
- Chern–Pontryagin Lagrangian, 927
- Chern–Simons action, 35
- Chern–Simons gauge theory, 35, 924, 1098, 1193, 1224
- Chern–Simons Lagrangian, 35, 925
- Chern–Simons Lagrangian density, 893, 926
- Chern–Simons theory, 1103
- chirality operator, 675
- Christoffel symbols, 20, 48, 60, 274, 290, 389, 434, 559, 834, 872, 936
- circle, 2
- circle bundle, 530
- circle group, 357
- classical Lie algebra, 770
- classical theory of Lie groups, 235
- Clebsch–Gordan condition, 715
- Clifford, 5
- Clifford action, 668, 726
- Clifford algebra, 666, 727
- Clifford algebras, 958
- Clifford module, 668
- closed form, 182
- closed manifold, 10
- closed string theories, 1140
- co–area formula, 1070
- cobasis, 179
- coboundary operator, 517
- cocycle condition, 489, 492
- codifferential, 191
- coframing, 179
- coherence laws, 133
- cohomology, 182
- cohomology class, 792
- cohomology theory, 508
- collision detection, 262
- commutative monoid, 508
- commutative ring, 14, 509
- commutator subalgebra, 653
- compact Abelian Lie group, 29
- compact manifold, 4
- compactness, 29
- complementary map, 803
- complex Clifford algebra, 958
- complex coordinates, 434
- complex Laplacian, 439
- complex manifold, 428
- complex phase–space manifold, 348
- complex structure, 429, 432
- complexified tangent space, 430
- composite bundle, 960
- composite connection, 814
- composite fibration, 814
- conatural projection, 154
- configuration bundle, 825, 830
- configuration manifold, 137, 143, 150, 801
- configuration space of sections, 801
- conformal field theory, 1145
- conformal invariance, 1183
- conformal Killing form, 442
- conformal Killing tensor–field, 443
- conformally flat gauge, 1183
- conjugation, 672
- connected components, 9
- connected Lie group, 246
- connected space, 9
- connectedness, 28
- connection, 16, 34, 467, 825
- connection form, 625
- connection homotopy, 278
- connection matrix, 455
- conservation law, 879
- constant of motion, 339

- constrained Hamiltonian equations, 901
- contact form, 880
- contact forms, 804
- contact manifold, 27, 880, 882
- contact map, 803
- contact transformation, 879
- contraction, 178, 670
- contravariant acceleration functor, 385
- control affine system, 538
- control vector-fields, 537
- controlled trajectory, 538
- controlled Van der Pol oscillator, 543
- coordinate 1-forms, 175
- coordinate ball, 145
- coordinate chart, 145
- coordinate domain, 145
- coordinate map, 145
- coproduct, 463
- correlation functions, 1118
- correspondence principle, 991
- cosmology, 974
- cotangent bundle, 25, 153
- cotangent space, 15, 153
- counit, 463
- coupling constant, 43
- coupling form, 626, 638
- covariant derivative, 467, 810
- covariant differential, 810, 838
- covariant differentiation, 273
- covariant force functor, 49, 384, 573, 1097
- covariant force law, 76, 77, 384, 573, 1097
- covariant forces, 290
- covariant formalism on smooth manifolds, 46
- covariant Hamiltonian, 900
- covariant Hamiltonian equations, 900
- cover, 2
- covering space, 487
- covertical connection, 809
- Coxeter graph, 243
- Coxeter number, 525
- Coxeter–Dynkin diagram, 243
- critical point, 312
- cross-section, 150
- cumulative distribution function, 985
- current conservation, 35
- curvature, 5
- curvature form, 512
- curvature of a connection, 811
- curvature operator, 278
- curvature-free connection, 812
- cuspy patterns, 1072
- CW-complex, 14
- D’Alambert–Lagrangian principle, 535
- D-branes, 36, 41
- Darboux’s Theorem, 459
- de Rham cohomology, 435, 508, 513, 636
- de Rham cohomology class, 336
- de Rham cohomology group, 182, 187, 844
- de Rham complex, 185
- de Rham Theorem, 186, 187
- deformed Einstein–Hilbert action, 472
- defuzzification, 586
- degree of non-holonomicity, 534
- derivation, 163
- deterministic chaos, 391
- diffeomorphism, 7, 29, 39, 148, 429
- diffeomorphism invariance, 707
- differentiability, 13
- differentiable manifold, 6, 12
- differentiable manifolds, 8
- differential system, 534
- dilaton, 44
- dilaton field, 1147
- dimension, 9
- Dimension Axiom, 508
- Dirac condition, 623
- Dirac constraint systems, 837
- Dirac equation, 442, 761
- Dirac Hamiltonian, 1170
- Dirac matrices, 200
- Dirac operator, 305, 726, 758
- Dirac quantization, 526, 614
- Dirac quantization rule, 87

- Dirac–Born–Infeld theory, 1239
- directional derivative, 29
- Dirichlet boundary conditions, 36
- Dirichlet branes, 1148
- disjoint union, 657
- dissipative structures, 983
- distribution, 473, 533
- distribution function, 984
- divergence of the stress tensor, 81
- divergence term, 879
- Dolbeault cohomology, 438
- Donaldson polynomials, 1118
- Donaldson theory, 757, 1103, 1118
- drift vector–field, 537
- dual Coxeter number, 776
- Duffing oscillator, 342, 347
- dynamical chaos, 1072
- dynamical connections, 827
- dynamical equation, 825
- dynamical intuition, 143
- Dynkin diagram, 240
- Dynkin index, 776
- edges, 5
- effective group action, 208
- Ehresmann connection, 572, 807
- Eilenberg–MacLane space, 521, 642
- Eilenberg–Steenrod Axioms, 508
- Einstein, 31
- Einstein equation, 87, 138, 142, 951, 965, 1056
- Einstein tensor, 22
- Einstein–Hilbert action, 32, 87, 300, 469, 972, 1056
- electro–weak interaction, 34
- elliptic Calogero–Moser system, 769
- elliptic geometry, 5
- emotion field, 603
- energy conservation law, 928
- energy function, 26
- energy functional, 317
- equal probabilities, 86
- equations of mathematical physics, 473
- equivariance condition, 531
- equivariant theory, 518
- Erlangen programme, 245
- Euclidean 3D space, 5
- Euclidean chart, 144
- Euclidean geometry, 1
- Euclidean image, 144
- Euclidean metric, 147, 979
- Euclidean nD space, 5
- Euclidean spaces, 2, 7, 8
- Euclidean triangulations, 1067
- Euclidean–Schwarzschild metric, 970
- Euler, 5
- Euler angles, 28
- Euler beta function, 40
- Euler character, 1118
- Euler characteristic, 5, 28, 271, 284, 1072, 1075, 1096, 1188
- Euler characteristics, 515
- Euler class, 514, 1117
- Euler number, 973
- Euler–Lagrange equations, 22
- Euler–Lagrangian equations, 84, 278, 879, 900, 1183
- Euler–Lagrangian functional derivative, 257
- Euler–Lagrangian operator, 842
- Euler–Poincaré characteristics, 188
- Euler–Poincaré equations, 291
- European option, 1026
- evolution operator, 165
- exact form, 182
- exact sequence, 508
- exponential, 672
- exponential law, 30
- exponential map, 29, 205, 248
- exponentiation of a vector–field, 247
- extended supersymmetry, 35
- extensive quantities, 473
- exterior algebra, 669
- exterior bundle, 493
- exterior calculus, 16
- exterior derivative, 180
- exterior differential, 609
- exterior differential forms, 174
- exterior differential system, 179, 879
- exterior horizontal form, 612
- exterior powers, 435

- exterior product, 669
- external configuration manifold, 47
- extraordinary cohomology theory, 508
- extremal, 25, 84
- extremal path, 84
- extrinsic view, 4
- faces, 5
- Faddeev–Popov procedure, 1107, 1126
- family of probability distributions, 333
- Faraday line, 706
- Faraday tensor, 124
- feature-space, 159
- feedback linearization, 539
- Fermi–Dirac statistics, 34
- Feynman, 31
- Feynman diagram, 1142
- Feynman kernel, 996
- Feynman path integral, 86, 983, 989, 994, 1079
- Feynman quantization, 85
- Feynman–Vernon formalism, 1041
- fibre, 150, 487
- fibre bundle, 15
- fibre-derivative, 258
- field strength, 1234
- fields, 84
- filtration, 670
- final state, 85
- finite-time probability distribution, 1033
- Finsler curvature tensor, 319
- Finsler energy function, 316
- Finsler information structure, 334
- Finsler manifold, 316
- Finsler metric, 334
- Finsler–Lagrangian field theory, 321
- first Cauchy law of motion, 81
- first integral, 339
- first superstring revolution, 41
- first variation, 878
- first variational formula, 841, 883, 884, 919
- first-order Lagrangian formalism, 801, 929, 957
- first-stage reducible, 1125
- fixed point, 340
- flag, 534
- flow, 172, 607
- flow line, 168
- flow property, 172
- flux class, 627
- flux group, 625
- flux homomorphism, 644
- Fock space, 527, 1148
- Fock state, 991
- Fokker–Planck, 87
- Fokker–Planck equation, 591, 988, 989, 1023
- foliation, 533, 553
- force equation, 342
- force-field psychodynamics, 1079
- forced Lagrangian equation, 535
- formal exponential, 173
- four fundamental forces, 33
- Fourier transform, 461
- Frölicher–Nijenhuis bracket, 613, 913
- frame bundle, 530
- Fredholm operator, 520, 1114, 1246
- free group action, 208
- free motion equation, 832
- free string, 1139
- Freed–Hopkins–Teleman Theorem, 526
- Frobenius Theorem, 553
- Frobenius–Cartan criterion, 263
- functional, 83
- functional manifold, 268
- Fundamental Theorem of calculus, 482
- Galilei group, 210
- gauge condition, 1014
- gauge field, 34, 527
- gauge form, 513
- gauge group, 30, 513, 731, 869
- gauge Lie group, 527
- gauge potentials, 921
- gauge theories, 34
- gauge transformation, 35, 490
- gauge transformations, 1234

- Gauss, 5
- Gauss map, 284
- Gauss–Bonnet formula, 271, 284, 1077
- Gauss–Bonnet Theorem, 188, 512
- Gauss–Kronecker curvature, 1077
- Gaussian curvature, 271, 1061
- general linear group, 211
- general linear Lie algebra, 196
- general principal vector-fields, 896
- general relativity, 292, 448
- general theory of systems, 393
- generalized function, 473
- generalized vector-field, 256
- generic energy condition, 966
- genus, 725, 1076
- geodesic, 17, 25, 170, 278
- geodesic deviation equation, 139, 140
- geodesic equation, 20, 278, 825
- geodesic flow, 290
- geodesic spray, 170, 290
- geometrical intuition, 144
- geometrical invariance group, 869
- geometrodynamical functor, 1082
- gerbe, 524
- ghost number, 1149
- global space-time hyperbolicity, 964
- graph, 485
- Gravity, 33
- Green–Gauss–Ostrogradsky Divergence Theorem, 480
- Green–Schwarz bosonic string theory, 1181
- Green–Schwarz mechanism, 41
- Gromov–Witten invariants, 632, 654
- Grothendieck Additivity Axiom, 515
- Grothendieck group, 508
- Grothendieck’s Riemann–Roch Theorem, 516
- group action, 98
- group homomorphism, 28
- group identity element, 203
- group inversion, 203
- group multiplication, 203
- group of rotations, 30
- group of symplectomorphisms, 625
- group orbit space, 208
- Haar measure, 207
- Hairy Ball Theorem, 513
- Hamel equations, 292
- Hamilton’s equations, 27
- Hamilton’s principle, 23
- Hamilton–de Donder equations, 843, 852
- Hamilton–Jacobi equation, 27
- Hamiltonian, 24, 26, 84
- Hamiltonian action, 372
- Hamiltonian automorphism, 628
- Hamiltonian bundle, 624
- Hamiltonian connection, 840, 844, 845, 931
- Hamiltonian conservation laws, 846
- Hamiltonian density, 84
- Hamiltonian dynamics, 154
- Hamiltonian energy function, 338
- Hamiltonian equations, 840
- Hamiltonian extension class, 628
- Hamiltonian flow, 27
- Hamiltonian form, 840, 841, 844, 845, 854, 900, 938, 946
- Hamiltonian jet-field, 938
- Hamiltonian map, 854
- Hamiltonian mechanical system, 338
- Hamiltonian mechanics, 13, 25, 26
- Hamiltonian structure, 626
- Hamiltonian vector-field, 26, 338
- harmonic forms, 436
- Hausdorff space, 145
- Hawking’s Euclidean quantum gravity, 302
- heat equation, 250, 261
- Heisenberg picture, 1007
- Heisenberg uncertainty relations, 87
- Heisenberg’s uncertainty principle, 41
- helicity, 502
- Hermitian inner product, 432, 686
- Hermitian metric, 432
- Hermitian operator, 85, 87
- Hermitian Riesz representation, 687, 688
- Hermitian structure, 686

- Hermiticity condition, 432
- Hessian, 276
- heuristic action paradigm, 83
- heuristic quantization rule, 711
- higher-order contact, 797
- higher-order tangency, 797
- Hilbert, 31
- Hilbert 5th problem, 236
- Hilbert action principle, 142
- Hilbert manifold, 147
- Hilbert space, 85, 87, 147, 992
- Hirzebruch's Riemann–Roch Theorem, 516
- Hodge decomposition, 436
- Hodge diamond, 441
- Hodge identities, 435
- Hodge numbers, 440, 1189
- Hodge star operator, 125, 178, 191, 681
- Hodge Theorem, 439, 520
- Hodge theory, 417
- Hodge's Theorem, 436
- holomorphic 1-form, 791
- holomorphic cotangent space, 431
- holomorphic function, 725
- holomorphic tangent space, 431
- holomorphicity constraints, 719
- holonomic atlas, 163
- holonomic coframes, 164
- holonomic connections, 827
- holonomic fibre bases, 486
- holonomic frames, 163
- holonomous frame field, 163
- holonomy, 125, 1195
- holonomy group, 738
- homeomorphism, 6–9, 429
- homoclinic orbits, 343
- homological algebra, 104, 510
- homology group, 105, 183
- homotopy fibration, 645
- homotopy lifting property, 486, 738
- homotopy operators, 186, 546
- Hopf algebra, 463
- Hopf–Rinow Theorem, 17
- horizontal density, 801, 883, 900, 957
- horizontal distribution, 812
- horizontal foliation, 812
- horizontal forms, 611
- horizontal splitting, 960
- human crowd, 603
- human–robot team, 154
- hyperbolic geometry, 5
- hyperkähler manifolds, 1187
- ideal
 - differential, 881
- Immirzi parameter, 299
- imprecision of measurement, 391
- independence condition, 881
- index, 655
- inertial metric tensor, 833
- infinite-dimensional neural network, 1080
- infinite-order jet space, 822
- infinitesimal generators, 29, 253
- information gain, 1018
- initial state, 85
- initial-date coordinates, 862
- inner product, 16
- inner product space, 992
- input vector-fields, 537
- insertion operator, 178
- integrable Hamiltonian system, 262
- integrable in the sense of Liouville, 771
- integrable system, 790
- integral curve, 168
- integral manifold, 533, 879, 881
- intensive quantities, 473
- interior product, 178, 670
- internal configuration manifold, 46
- intertwining tensor, 714
- intrinsic definition for differentiable manifolds, 6
- intrinsic view, 4
- invariant interval, 1173
- invariant of Poincaré–Cartan, 841
- invariant symplectic form, 690
- invariant tori, 343
- involution closure, 552
- involution distribution, 533
- irreducible representation, 527

- isometric, 17
- isomorphic, 28
- isotropy group, 208
- Itô lemma, 1026
- iterated fibration of Kähler manifolds, 653
- Ito prescription, 1019
- Ito stochastic integral, 987
- Jacobi equation of geodesic deviation, 280
- Jacobi fields, 280, 313
- Jacobi identity, 126, 202, 362
- Jacobi operator, 879
- jet, 798, 880
- jet bundle, 16, 801, 803
- jet field, 807
- jet functor, 804
- jet space, 251, 275, 797
- Jones polynomial, 1195
- K-theory, 494, 508, 515
- Kähler form, 740
- Kähler manifold, 733
- Kähler metric, 742
- Kähler potential, 733, 736
- Kähler condition, 417
- Kähler form, 432
- Kähler gauge, 1225
- Kähler identities, 435
- Kähler manifold, 27, 416, 433
- Kähler metric, 433
- Kähler potential, 434, 724, 1227
- Kähler structure, 433
- Kählerity condition, 433
- Kaluza–Klein theory, 44
- Kepler problem, 27
- Killing equation, 441
- Killing form, 246, 354, 785
- Killing spinor–field, 442
- Killing tensor–field, 443, 447
- Killing vector–field, 441, 446
- Killing–Riemannian geometry, 441
- Killing–Yano equation, 442, 450
- Killing–Yano tensor, 450
- Klein bottle, 487
- Klein–Gordon Lagrangian, 1013
- Kodaira Embedding Theorem, 435
- Korteweg–de Vries equation, 262, 598
- Kostant–Souriau prequantization, 622
- Kuiper’s Theorem, 521
- Lagrange multipliers, 1179
- Lagrange’s equations, 22
- Lagrange–Poincaré equations, 292
- Lagrangian, 15, 84, 289, 877
- Lagrangian constraint space, 841
- Lagrangian density, 84, 801, 888, 1013, 1085, 1175
- Lagrangian dynamics, 153
- Lagrangian mechanics, 13, 22, 26
- Lagrangian submanifold, 28
- Landau pole, 736
- Langevin, 87
- Laplace equation, 252
- Laplace–Beltrami operator, 191, 277
- Laplacian operator, 482
- Laplacian symmetry, 445
- large, 646
- laws of motion, 166
- Lax equation, 771, 780
- Lax operator, 790
- Lax pair equation, 450
- Lax tensor, 450
- Lax type representation, 262
- leaf space, 553
- Lebesgue measure, 207
- Lefschetz condition, 637
- Lefschetz Theorem, 436
- left ideal, 202
- left-invariant Lagrangian, 291
- left-invariant Riemannian metric, 354
- Legendre bundle, 900
- Legendre map, 258, 370, 841, 843, 885
- Legendre submanifold, 881
 - transverse, 881, 883
- Legendre transformation, 24
- Leibniz rule, 181, 464
- Lepagean equivalent, 842
- Leray cohomology spectral sequence, 665
- Leray–Hirsch basis., 663

- Leray–Serre cohomology spectral sequence, 638
- level vector, 775
- Levi–Civita connection, 19, 48, 273, 379, 416, 434, 450, 451, 502, 535, 559, 572, 728, 957
- Levi–Civita symbol, 1177
- Lewinian force–field theory, 1081
- Lewinian psychodynamics, 602
- Lie, 6, 28
- Lie algebra, 27, 29, 196, 362
- Lie algebra homomorphism, 202
- Lie algebra simple root, 242
- Lie bracket, 19, 29, 194, 463, 606
- Lie bracket property, 256
- Lie derivative, 29, 163, 192, 338, 446, 610, 837, 838, 840
- Lie Functor, 30
- Lie functor, 205
- Lie group, 4, 203
- Lie product, 553
- Lie structural constants, 362
- Lie subalgebra, 202
- Lie super–algebra, 671, 1105
- Lie–derivative neuro–classifier, 200
- Lie–group–homomorphism, 28
- Lie–invariant geometric objects, 263
- Lie–Lagrangian biodynamical functor, 385
- Lie–Poisson bracket, 361
- Lie–Poisson neuro–classifier, 383
- lifted action, 373
- limit set, 341
- line bundle, 514, 524, 691
- line bundles, 494
- linear connection, 807
- linear controllability, 544
- linear isotopy, 648
- linearized Hamiltonian dynamics, 1073
- Liouville equation, 27, 988
- Liouville integrability, 451
- Liouville measure, 25
- Liouville operator, 192
- Liouville Theorem, 363
- Liouville’s Theorem, 27
- Liouville–Arnold Theorem, 367
- Lipschitz condition, 171
- local connection form, 815
- local coordinate system, 6
- local geodesics, 137
- locally accessible system, 556
- locally topologically equivalent, 137
- loop algebra, 706
- loop quantum gravity, 527, 702
- loop representation, 717
- loop variables, 297
- Lorentz metric tensor, 1181
- Lorentz–invariant theories, 1140
- Lorentzian dynamical triangulations, 1059
- Lorentzian–de Sitter metric, 979
- Lorenz dynamics, 506
- Lorenz flow, 506
- low energy effective action, 1119
- Lyapunov exponent, 361, 1072
- Lyapunov stable, 865, 866
- Möbius strip, 487
- manifold, 1, 8, 137, 143
- manifold structure, 145
- manifold with boundary, 186
- manifoldness, 5
- marked symplectic manifold, 630
- Markov assumption, 988
- Markov chain, 985
- Markov stochastic process, 985, 1083
- Markov–Gaussian stochastic systems, 87
- mass conservation principle, 78
- Master equation, 988
- Mathai–Quillen formula, 1115
- mathematical induction, 6
- matrix representation, 239
- Maupertius action principle, 277
- Maurer–Cartan connection, 1117
- Maurer–Cartan equations, 263
- maximal atlas, 7
- maximal geodesic, 170
- maximal integral curve, 168
- Maxwell equations, 33, 124
- Maxwell gauge field theory, 34

- Maxwell Lagrangian, 34
- Maxwell–Dirac duality, 724
- Mayer–Vietoris sequence, 660
- Mayer–Vietoris Theorem, 508
- mean curvature, 877
- mean–field theory, 1075
- Melnikov function, 348
- mental force law, 1097
- meromorphic 1–form, 725, 736
- mesoscopic order parameters, 1016
- metric, 468
- metric manifold, 705
- metric SEM–tensor, 924, 930, 936, 1198
- metric space, 17
- metric tensor, 20
- midpoint discretization, 1019
- mirror symmetry, 1189
- model space, 147
- module, 1242
- moduli space, 35, 524, 720, 724
- momentum constraint equation, 977
- momentum map, 354, 372, 446
- momentum phase–space manifold, 258, 340, 387, 837, 839, 844, 848
- monodromy, 738
- monodromy group, 738
- monoid, 129
- monoidal category, 129
- Montonen–Olive duality, 719, 1119
- Montonen–Olive mass formula, 722
- Morita–equivalent, 1250
- morphism of vector–fields, 173
- Morse function, 312, 1073, 1077
- Morse lemma, 1073
- Morse numbers, 1078
- Morse theory, 311, 1073, 1077, 1098
- Moser’s homotopy argument, 632
- motion planning, 262
- motivation–behavior conservation law, 606
- motivational factor manifold, 603
- motivational factor–structure, 603
- Moyal bracket, 1233
- Moyal product, 458, 470
- multi–index, 250, 879
- multiindex, 177
- multimomentum Hamiltonian, 938
- multivector–field, 608
- Nambu–Goto action, 1145, 1172, 1183
- natural geometrical structures, 128
- natural projection, 152
- Navier–Stokes equations, 83
- Neumann boundary condition, 1149
- neural path integral, 1045
- Neveu–Schwarz B–field, 1235, 1243
- Newman–Penrose equation, 965
- Newton’s Second Law, 23
- Newtonian equation of motion, 166
- Newtonian fluid, 82
- Newtonian mechanics, 22
- Nijenhuis differential, 614, 913
- Noether conservation law weak, 923
- Noether conservation laws, 446
- Noether current, 923, 925
- Noether identities, 923
- Noether Theorem, 24, 879
- Noether–Lagrangian symmetry, 257
- Noetherian ring, 100
- non–Abelian field strength, 1191
- non–equilibrium nonlinear multivariate, 1016
- Non–Euclidean geometry, 5
- noncommutative coordinates, 459
- noncommutative gravity theory, 463
- noncommutative phase–space, 459
- noncommutative product, 215
- noncommutative Yang–Mills theory, 1234
- nondegenerate quadratic forms, 958
- nonholonomic connection, 536
- nonholonomic coordinates, 275
- nonlinear control system, 537
- nonlinear control theory, 1141
- nonlinear controllability criterion, 551
- nonlinear factor analysis, 603
- nonlinear MIMO–systems, 537
- nonlinear Schrödinger equation, 597
- nonlinear sigma model, 1145
- normal bundle, 490

- normal vector-field, 167
- obstruction theory, 642
- on-shell reducible, 1126
- one-parameter group of
 - diffeomorphisms, 172, 341
- open manifold, 10
- open string theories, 1140
- operator algebras, 511
- orbifold, 13, 1187
- order parameter equations, 1016
- oriented strings, 1141
- orthogonal group, 30
- overlap, 7
- parabolic Einstein equation, 285
- parallel transport, 20, 139, 290, 379, 423, 527
- parameter-space of probability
 - distributions, 333
- particles, 84
- partition function, 1070, 1097, 1194
- path, 9
- path integral, 86, 970, 971, 1020
- path measure, 86, 1017
- path-connected space, 9
- path-integral expression, 1011
- path-integral formalism, 1008
- path-integral formulation, 24, 1007
- path-integral quantization, 1007
- path-ordered exponential, 527
- Pauli exclusion principle, 34
- perturbation, 656
- perturbation theory, 345
- perturbative path integral, 1055
- perturbative string theory, 1146
- Peyrard-Bishop system, 1074
- Pfaff Theorem, 880
- Pfaffian forms, 608
- Pfaffian system, 387, 533, 879
- phase, 991
- phase space, 5
- phase trajectory, 340
- phase transition, 1069, 1074
- phase-flow, 341
- phase-space, 25
- phase-space path integral, 1003
- Philip Hall basis, 553
- Picard groupoid, 524
- Planck constant, 459
- Planck length, 36, 1139
- Poincaré, 6
- Poincaré conjecture, 6, 150
- Poincaré duality, 440
- Poincaré lemma, 73, 74, 182
- Poincaré maps, 346
- Poincaré-Cartan form, 842, 882
- Poincaré-Hopf Theorem, 188
- Poincaré algebra, 462
- Poincaré dual, 528
- Poincaré duality, 516
- Poincaré-Birkhoff-Witt property, 460
- point evaluation map, 635
- Poisson bivector-field, 609
- Poisson bracket, 26, 27, 84, 260, 339, 375, 451, 837, 907
- Poisson detection statistics, 992
- Poisson evolution equation, 361
- Poisson manifold, 27, 361
- Poisson structure, 459
- Poisson tensor-field, 369
- polarization, 676
- Polyakov action, 40, 1145
- Polyakov equation, 45
- polysymplectic phase-space, 840
- Pontryagin class, 512, 1121
- Pontryagin Maximum Principle, 570, 1020
- Ponzano-Regge ansatz, 301
- Ponzano-Regge quantization, 302
- positive chirality condition, 1122
- prepoint discretization, 1019
- prepotential, 722, 725
- presymplectic Hamiltonian systems, 837
- principal bundle, 488, 529
- principal connections, 815
- Principle of Democracy, 975
- Principle of stationary action, 24
- probability amplitude, 85, 1007, 1097
- probability distance, 333
- probability divergence, 333

- probability manifold, 333
- product manifold, 1123
- projectable vector-field, 612, 823
- projected connection, 536
- projective bundles, 517
- prolongation, 250, 502
- prolonged group action, 253
- propagator, 86, 996
- pull-back, 164
- pull-back vector bundle, 492
- pull-back-valued forms, 613
- push-forward, 164

- qualitative ODE theory, 360
- quantization, 85
- quantum algebra, 619
- quantum brain modelling, 1079
- quantum bundle, 622
- quantum chromodynamics, 34, 40
- quantum coherent state, 991
- quantum evolution pictures, 85
- quantum field theory, 34, 35, 292, 527, 1097
- quantum fields, 86
- quantum gauge theory, 722
- quantum geometry, 1180
- quantum gravity, 38, 1055
- quantum Hilbert space, 520
- quantum measure, 1193
- quantum operator, 716
- Quantum superposition, 85
- quantum-mechanical particles, 86
- quaternions, 215
- Quillen's plus construction, 511
- Quillen-Suslin Theorem, 510
- quotient representation, 240

- Ramond-Ramond field, 526
- random variable, 984
- random walk, 985
- rank condition, 548
- ray bundle, 692
- Ray-Singer torsion, 1118
- reachable sets, 538
- real Lie group, 28
- reduced curvature 1-form, 270
- reduced phase-space, 373
- reducible constraints, 854
- Regge calculus, 1061
- Regge geometries, 1057
- Regge simplicial action, 1063
- regular divergence, 334
- regularization, 655
- Reidemeister moves, 132
- related vector-fields, 164
- relative degree, 542
- relativistic mechanics, 618
- repeated jet space, 819
- representation of a Lie group, 239
- representative point, 143
- Rho-tensor, 694
- Ricci flow, 285
- Ricci scalar curvature, 22
- Ricci tensor, 22, 140, 272, 278, 435
- Riemann, 5
- Riemann curvature tensor, 21, 56, 71, 139, 141, 271, 278, 835, 950, 1095
- Riemann sphere, 18, 510, 513
- Riemann surface, 17, 433, 725
- Riemann tensor, 435
- Riemann-Roch Theorem, 515
- Riemannian geometry, 19
- Riemannian manifold, 16, 166, 178
- Riemannian manifolds, 5
- Riemannian metric, 379
- Riemannian metric tensor, 271
- right translation, 204
- rigid body with a fixed point, 354
- root system, 240
- rotation group, 28

- S-duality, 44
- saddle point approximation, 978
- Sasakian metric, 322, 504
- scalar curvature, 272, 279
- scalar Gaussian curvature, 281
- scalar-field, 79
- scattering, 976
- Schouten-Nijenhuis bracket, 608
- Schrödinger equation, 991, 1002
- Schrödinger operators, 619
- Schrödinger picture, 1007

- Schrödinger quantization, 616
- Schwarz-type TQFT, 1121
- Schwarzschild metric form, 970
- second superstring revolution, 44
- second tangent bundle, 495
- second tangent maps, 156
- second variation, 879
- second variation formula, 281
- second-countable space, 145
- sectional curvature, 271
- Seiberg–Witten gauge theory, 35
- self-dual basis, 681
- SEM conservation laws, 935
- SEM-tensors
 - Hamiltonian, 941
 - Lagrangian, 933
- semisimple representation, 246
- separatrix, 343
- Serre fibration, 645
- Serre's Conjecture, 510
- set, 116
- shape operator, 1077
- shear viscosity, 82
- sheaves, 14
- signature, 958
- simple Lie group, 245
- simplicial set, 135
- Sine–Gordon equation, 598
- sliding filament theory of muscular
 - contraction, 381
- small-time local controllability, 556
- small-time locally controllable, 538
- smooth homomorphism, 203
- smooth manifolds, 4, 8
- smoothness, 7
- Sobolev norm, 521
- Sobolev–Schwartz functional analysis, 473
- soldering curvature, 811
- soldering form, 612
- space of paths, 86
- space of quantum states, 523
- space-time, 5
- space-time manifold, 137
- special Lagrangian, 437
- sphere, 3
- sphere bundle, 487
- spin, 34
- spin connection, 708
- spin connection 1-form, 1122
- spin foam models, 308
- Spin group, 672
- spin network, 713
- spin network theory, 714
- spin networks, 1056
- Spin-connection, 728
- spinor, 200
- Spinor bundle, 725
- spinor bundle, 727
- spinor contraction, 691
- spinor Lie group, 956
- spinor module, 676
- spinor-twistor object, 699
- stable equivalence, 509
- stack, 524
- Standard Model, 30, 33, 292, 702
- state feedback, 555
- state manifold, 537
- static gauge, 1182
- Steenrod operation, 522
- Stiefel–Whitney class, 512
- stochastic forces, 391
- stochastic integral, 986
- Stoke's Theorem, 529
- Stokes fluid, 82
- Stokes formula, 187
- Stokes Theorem, 480, 514
- strange homology group, 631
- Stratonovich prescription, 1019
- string corrections, 1140
- string cosmology, 296
- string coupling constant, 44
- string length, 1172
- string tension, 1145
- string-field, 1224
- string-field-theory action, 1149
- strong energy condition, 965
- structural group, 625
- structure equations, 287
- submanifold, 28
- super-algebra, 667
- super-field, 1215

- super-space, 1215
- supercovariant derivative, 743
- superstring theory, 35, 1146
- supersymmetry, 33, 38, 1146
- support of a vector-field, 172
- surface forces, 80
- surface of Earth, 1
- surgery theory, 511
- surjection lemma, 651
- SW-spectral curve, 775
- symbol map, 670
- symmetric affine connection, 273
- symmetric Weyl ordering, 460
- symmetry, 879
- symmetry group, 250
- symplectic bundle, 626
- symplectic condition, 666
- symplectic connection form, 649
- symplectic fibre bundle, 625
- symplectic foliation, 843
- symplectic form, 335, 610, 689
- symplectic group, 30, 335
- symplectic manifold, 15, 25, 336, 338
- symplectic manifolds, 5
- symplectic map, 336
- symplectic potential, 615
- symplectic Riesz representation, 689, 690
- symplectic structure, 689
- symplectic volume form, 25
- symplectomorphism, 27, 335
- synthetic differential geometry, 473
- system variables, 83
- T-duality, 43
- tachyon field, 1147
- tangent bundle, 14, 17, 152
- tangent dynamics equation, 1072
- tangent map, 151, 152
- tangent space, 15, 16, 150
- tangent vector-field, 167
- tangent-valued r -forms, 611
- tangent-valued horizontal form, 612
- tangle, 129
- tangle diagram, 131
- tautological bundle, 510
- tautological section, 692
- Taylor's Theorem, 12
- tensor, 15, 16
- tensor algebra, 667
- tensor bundle, 15, 163, 498
- tensor product of vector bundles, 509
- tensor-field, 163
- tensor-product connection, 812
- tetrad gravitational field, 955
- the twisted elliptic CM-systems, 771
- thermodynamic partition function, 971
- three-body problem, 357
- time-dependent flow, 166
- time-dependent mechanics, 830
- time-dependent state of a quantum system, 85
- time-dependent vector-field, 168, 173
- Toda chain, 790
- Toda molecule, 353
- topological Bogomol'nyi action, 1125
- topological group, 203
- topological hypothesis, 1070
- topological invariant, 5, 1076, 1121, 1122
- topological K-theory, 508, 510
- topological manifold, 4, 6, 8
- topological operads, 134
- topological property, 5
- topological quantum field theory, 1097
- topological space, 8
- topological structure, 6
- topological Theorem, 1072
- torsion, 19
- torsion of a connection, 811
- torsion tensor, 536
- torus, 3
- total Hamiltonian, 1176
- transformation
 - contact, 878
 - point, 878
- transformation classical, 878
- transformation gauge, 878
- transformation of coordinates, 7
- transition, 85

- transition amplitude, 993, 1001
- transition function, 7
- transition functions, 144, 489
- transition map, 2, 7
- transition probability, 993
- transition probability distribution, 1031
- transitive group action, 208
- transversal bundle, 533
- tricategories, 134
- trivial bundle, 15
- trivial fibration, 487
- trivial flux, 628
- twisted, 773, 782, 783
- twisted Dirac equation, 1122
- twisted K-groups, 517
- twisted K-theory, 515
- twistor bundle, 695
- twistor equation, 442
- twistors, 691

- unified field theory, 31
- universal bundle, 653

- vacuum state, 992
- vector bundle, 487, 491
- velocity equation, 342
- velocity phase-space manifold, 150, 258, 386, 824, 830
- velocity vector-field, 150
- vertical bundle, 495
- vertical connection, 809
- vertical cotangent bundle, 499
- vertical covariant differential, 815
- vertical lift, 495
- vertical tangent bundle, 499
- vertical vector-field, 612
- vertical-valued horizontal form, 612
- vertices, 5
- vierbein, 468
- Virasoro operators, 1186
- visual physical intuition, 32
- volatility, 1026
- volume element, 1174
- volume forces, 80
- volume form, 191

- volume viscosity, 82
- von Neumann dimension theory, 518

- Wang differential, 665
- Wang sequence, 664
- wave psi-function, 85
- wave-particle duality, 991
- weak conservation law, 893
- weak energy condition, 457, 965
- wedge product, 177
- weighted function, 691
- Weitzenböck formula, 728
- Weyl, 6
- Weyl deformation quantization, 459
- Weyl group, 732
- Weyl homomorphism, 927
- Weyl invariance, 1183
- Weyl spinor, 721, 1122
- Weyl tensor, 141
- Weyl vector, 776
- Wheeler-DeWitt equation, 977
- Whitham deformations, 725
- Whitney, 6
- Wick rotation, 970, 1065, 1152
- Wiener process, 988
- Wigner function, 1040
- Wilson line, 527, 1234
- Wilson loop, 527, 1118, 1195
- winding number, 350
- Wirtinger's inequality, 435
- Witten, 31, 1120
- Witten's TQFT, 1098
- world-sheet, 1139
- world-sheet dynamics, 1181
- world-volume, 36

- Yang-Baxter equation, 1103
- Yang-Lee Theorem, 1069
- Yang-Mills action, 39
- Yang-Mills covariant derivative, 126
- Yang-Mills gauge field theory, 34
- Yang-Mills gauge theory, 13, 899, 1053, 1099
- Yang-Mills Lagrangian, 893, 915, 920, 921
- Yang-Mills relation, 126

Yang–Mills theory, 39, 297
 higher, 127

Zamolodchikov metric, 1220

Zorn’s lemma, 18